(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
22 January 2004 (22.01.2004)

PCT

(10) International Publication Number
WO 2004/006838 A2

(51) International Patent Classification⁷: A61K

(21) International Application Number:
PCT/US2003/021730

(22) International Filing Date: 15 July 2003 (15.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/395,632   15 July 2002 (15.07.2002)   US

(71) Applicant (for all designated States except US): SUGEN, INC. [US/US]; 230 East Grand Avenue, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WHYTE, David [US/US]; 2623 Barclay Way, Belmont, CA 94002 (US). MANNING, Gerard [IE/US]; 844 Fremont Street, #4, Menlo Park, CA 94025 (US). CAENEPEEL, Sean [US/US]; 1755 Trinity Avenue, Apt. 13, Walnut Creek, CA 94596 (US).

(74) Agents: BURROUS, Beth, A. et al.; Foley & Lardner, Washington Harbour, 3000 K Street N.W., Suite 500, Washington, DC 20007 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL KINASES

(57) Abstract: The present invention relates to kinase polypeptides, nucleotide sequences encoding the kinase polypeptides, as well as various products and methods useful for the diagnosis and treatment of various kinase-related diseases and conditions. Through the use of a bioinformatics strategy, mammalian members of the of PTK's and STK's have been identified and their protein structure predicted.

NOVEL KINASES

This application claims priority to U.S. Provisional Application No. 60/395,632, which was filed on July 15, 2002.

FIELD OF THE INVENTION

The present invention relates to kinase polypeptides, nucleotide sequences encoding the kinase polypeptides, as well as various products and methods useful for the diagnosis and treatment of various kinase-related diseases and conditions.

BACKGROUND OF THE INVENTION

The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to be or to describe prior art to the invention.

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of proteins, which enables regulation of the activity of mature proteins by altering their structure and function.

Protein phosphorylation plays a pivotal role in cellular signal transduction. Among the biological functions controlled by this type of postranslational modification are: cell division, differentiation and death (apoptosis); cell motility and cytoskeletal structure; control of DNA replication, transcription, splicing and translation; protein translocation events from the endoplasmic reticulum and Golgi apparatus to the membrane and extracellular space; protein nuclear import and export; regulation of metabolic reactions, etc. Abnormal protein phosphorylation is widely recognized to

be causally linked to the etiology of many diseases including cancer as well as immunologic, neuronal and metabolic disorders.

The following abbreviations are used for kinases throughout this application:

ASK	Apoptosis signal-regulating kinase
CaMK	Ca2+/calmodulin-dependent protein kinase
CCRK	Cell cycle-related kinase
CDK	Cyclin-dependent kinase
CK	Casein kinase
DAPK	Death-associated protein kinase
DM	myotonic dystrophy kinase
Dyrk	dual-specificity-tyrosine phosphorylating-regulated kinase
GAK	Cyclin G-associated kinase
GRK	G-protein coupled receptor
GuC	Guanylate cyclase
HIPK	Homeodomain-interacting protein kinase
IRAK	Interleukin-1 receptor-associated kinase
MAPK	Mitogen activated protein kinase
MAST	Microtubule-associated STK
MLCK	Myosin-light chain kinase
MLK	Mixed lineage kinase
NEK	NimA-related protein kinase (=NEK)
PKA	cAMP-dependent protein kinase
RSK	Ribosomal protein S6 kinase
RTK	Receptor tyrosine kinase
SGK	Serum and glucocorticoid-regulated kinase
STK	serine threonine kinase
ULK	UNC-51-like kinase

Protein kinases in eukaryotes phosphorylate proteins on the hydroxyl substituent of serine, threonine and tyrosine residues, which are the most common phospho-acceptor

amino acid residues. However, phosphorylation on histidine has also been observed in bacteria.

The presence of a phosphate moiety modulates protein function in multiple ways. A common mechanism includes changes in the catalytic properties (Vmax and Km) of an enzyme, leading to its activation or inactivation.

A second widely recognized mechanism involves promoting protein-protein interactions. An example of this is the tyrosine autophosphorylation of the ligand-activated EGF receptor tyrosine kinase. This event triggers the high-affinity binding to the phosphotyrosine residue on the receptor's C-terminal intracellular domain of the SH2 motif of the adaptor molecule Grb2. Grb2, in turn, binds through its SH3 motif to a second adaptor molecule, such as SHC. The formation of this ternary complex activates the signaling events that are responsible for the biological effects of EGF. Serine and threonine phosphorylation events also have been recently recognized to exert their biological function through protein-protein interaction events that are mediated by the high-affinity binding of phosphoserine and phosphothreonine to WW motifs present in a large variety of proteins (Lu, P.J. et al (1999) Science 283: 1325-1328).

A third important outcome of protein phosphorylation is changes in the subcellular localization of the substrate. As an example, nuclear import and export events in a large diversity of proteins are regulated by protein phosphorylation (Drier E.A. et al (1999) Genes Dev 13: 556-568).

Protein kinases are one of the largest families of eukaryotic proteins with several hundred known members. These proteins share a 250-300 amino acid domain that can be subdivided into 12 distinct subdomains that comprise the common catalytic core structure. These conserved protein motifs have recently been exploited using PCR-based and bioinformatic strategies leading to a significant expansion of the known kinases.

Kinases largely fall into two groups: those specific for phosphorylating serines and threonines, and those specific for phosphorylating tyrosines. Some kinases, referred

-3-

to as "dual specificity" kinases, are able to phosphorylate tyrosine as well as serine/threonine residues.

Protein kinases can also be characterized by their location within the cell. Some kinases are transmembrane receptor-type proteins capable of directly altering their catalytic activity in response to the external environment such as the binding of a ligand. Others are non-receptor-type proteins lacking any transmembrane domain. They can be found in a variety of cellular compartments from the inner surface of the cell membrane to the nucleus.

Many kinases are involved in regulatory cascades wherein their substrates may include other kinases whose activities are regulated by their phosphorylation state. Ultimately the activity of some downstream effector is modulated by phosphorylation resulting from activation of such a pathway. The conserved protein motifs of these kinases have recently been exploited using PCR-based cloning strategies leading to a significant expansion of the known kinases.

Multiple alignment of the sequences in the catalytic domain of protein kinases and subsequent parsimony analysis permits the segregation of related kinases into distinct branches of subfamilies including: tyrosine kinases (PTK's), dual-specificity kinases, and serine/threonine kinases (STK's). The latter subfamily includes cyclic-nucleotide-dependent kinases, calcium/calmodulin kinases, cyclin-dependent kinases (CDK's), MAP-kinases, serine-threonine kinase receptors, and several other less defined subfamilies.

The protein kinases may be classified into several major groups including AGC, CAMK, Casein kinase 1, CMGC, STE, tyrosine kinases, and atypical kinases (Plowman, GD *et al.*, *Proceedings of the National Academy of Sciences*, USA, Vol. 96, Issue 24, 13603-13610, November 23, 1999; see also www.kinase.com). Within each group are several distinct families of more closely related kinases. In addition, there is a group designated "other" to represent several smaller families. In addition, an "atypical" family represents those protein kinases whose catalytic domain has little

or no primary sequence homology to conventional kinases, including the alpha
kinases, pyruvate dehydrogenase kinases, A6 kinases and PI3 kinases.

AGC group

The AGC kinases are basic amino acid-directed enzymes that phosphorylate residues
found proximal to Arg and Lys. Examples of this group are the G protein-coupled
receptor kinases (GRKs), the cyclic nucleotide-dependent kinases (PKA, PKC, PKG),
NDR or DBF2 kinases, ribosomal S6 kinases, AKT kinases, myotonic dystrophy
kinases (DMPKs), MAPK interacting kinases (MNKs), MAST kinases, and the
YANK family.

GRKs regulate signaling from heterotrimeric guanine protein coupled receptors
(GPCRs). Mutations in GPCRs cause a number of human diseases, including retinitis
pigmentosa, stationary night blindness, color blindness, hyperfunctioning thyroid
adenomas, familial precocious puberty, familial hypocalciuric hypercalcemia and
neonatal severe hyperparathroidism (OMIM, http: //www.ncbi.nlm.nih.gov/Omim/).
The regulation of GPCRs by GRKs indirectly implicates GRKs in these diseases.

The cAMP-dependent protein kinases (PKA) consist of heterotetramers comprised of
2 catalytic (C) and 2 regulatory (R) subunits, in which the R subunits bind to the
second messenger cAMP, leading to dissociation of the active C subunits from the
complex. Many of these kinases respond to second messengers such as cAMP
resulting in a wide range of cellular responses to hormones and neurotransmitters.

AKT is a mammalian proto-oncoprotein regulated by phosphatidylinositol 3-kinase
(PI3-K), which appears to function as a cell survival signal to protect cells from
apoptosis. Insulin receptor, RAS, PI3-K, and PDK1 all act as upstream activators of
AKT, whereas the lipid phosphatase PTEN functions as a negative regulator of the
PI3-K/AKT pathway. Downstream targets for AKT-mediated cell survival include
the pro-apoptotic factors BAD and Caspase9, and transcription factors in the forkhead
family, such as DAF-16 in the worm. AKT is also an essential mediator in insulin
signaling, in part due to its use of GSK-3 as another downstream target.

-5-

The S6 kinases (RSK) regulate a wide array of cellular processes involved in mitogenic response including protein synthesis, translation of specific mRNA species, and cell cycle progression from G1 to S phase. One of the RSK genes has been localized to chromosomal region 17q23 and is amplified in breast cancer (Couch, *et al.*, Cancer Res. 1999 Apr 1;59(7): 1408-11).

CAMK Group

The CAMK kinases are also basic amino acid-directed kinases. They include the Ca2+/calmodulin-regulated and AMP-dependent protein kinases (AMPK), myosin light chain kinases (MLCK), MAP kinase activating protein kinases (MAPKAPKs), checkpoint 2 kinases (CHK2), death-associated protein kinases (DAPKs), phosphorylase kinase (PHK), Rac and Rho-binding Trio kinases, a "unique" family of CAMKs, and the MARK family of protein kinases.

The MARK family of STKs are involved in the control of cell polarity, microtubule stability and cancer. One member of the MARK family, C-TAK1, has been reported to control entry into mitosis by activating Cdc25C which in turn dephosphorylates Cdc2.

CMGC Group

The CMGC kinases are "proline-directed" enzymes phosphorylating residues that exist in a proline-rich context. They include the cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), GSK3s, RCKs, (dual-specific tyrosine kinases) DYRKs, (SR-protein specific kinase) SRPKs, and CLKs. Most CMGC kinases have larger-than-average kinase domains owing to the presence of insertions within subdomains X and XI.

CDKs play a pivotal role in the regulation of mitosis during cell division. The process of cell division occurs in four stages: S phase, the period during which chromosomes duplicate, G2, mitosis and G1 or interphase. During mitosis the duplicated chromosomes are evenly segregated allowing each daughter cell to receive a complete copy of the genome. A key mitotic regulator in all eukaryotic cells is the STK cdc2, a

-6-

CDK regulated by cyclin B. However some CDK-like kinases, such as CDK5 are not cyclin associated nor are they cell cycle regulated.

MAPKs play a pivotal role in many cellular signaling pathways, including stress response and mitogenesis (Lewis, T. S., Shapiro, P. S., and Ahn, N. G. (1998) Adv. Cancer Res. 74, 49-139). MAP kinases can be activated by growth factors such as EGF, and cytokines such as TNF-alpha. In response to EGF, Ras becomes activated and recruits Raf1 to the membrane where Raf1 is activated by mechanisms that may involve phosphorylation and conformational changes (Morrison, D. K., and Cutler, R. E. (1997) *Curr. Opin. Cell Biol.* 9, 174-179). Active Raf1 phosphorylates MEK1 which in turn phosphorylates and activates the ERKs subfamily of MAPKs. DYRKS are dual-specificity tyrosine kinases.

Tyrosine Protein Kinase Group

The tyrosine kinase group encompass both cytoplasmic (*e.g. src*) as well as transmembrane receptor tyrosine kinases (*e.g.* EGF receptor). These kinases play a pivotal role in the signal transduction processes that mediate cell proliferation, differentiation and apoptosis.

STE Group

The STE family refers to the 3 classes of protein kinases that lie sequentially upstream of the MAPKs. This group includes STE7 (MEK or MAP2K) kinases, STE11 (MEKK or MAP2K) kinases and STE20 (MEKKK or MAP4K) kinases. In humans, several protein kinase families that bear only distant homology with the STE11 family also operate at the level of MAP3Ks including RAF, MLK, TAK1, and COT. Since crosstalk takes place between protein kinases functioning at different levels of the MAPK cascade, the large number of STE family kinases could translate into an enormous potential for upstream signal specificity. This also includes homologues of the yeast sterile family kinases (STE), which refers to 3 classes of kinases which lie sequentially upstream of the MAPKs;

The prototype STE20 from baker's yeast is regulated by a hormone receptor, signaling to directly affect cell cycle progression through modulation of CDK activity. It also

coordinately regulates changes in the cytoskeleton and in transcriptional programs in a bifurcating pathway. In a similar way, the homologous kinases in humans are likely to play a role in extracellular regulation of growth, cell adhesion and migration, and changes in transcriptional programs, all three of which have critical roles in tumorigenesis. Mammalian STE20-related protein kinases have been implicated in response to growth factors or cytokines, oxidative-, UV-, or irradiation-related stress pathways, inflammatory signals (*e.g.* TNFα), apoptotic stimuli (*e.g.* Fas), T and B cell costimulation, the control of cytoskeletal architecture, and cellular transformation. Typically the STE20-related kinases serve as upstream regulators of MAPK cascades. Examples include: HPK1, a protein-serine/threonine kinase (STK) that possesses a STE20-like kinase domain that activates a protein kinase pathway leading to the stress-activated protein kinase SAPK/JNK; PAK1, an STK with an upstream GTPase-binding domain that interacts with Rac and plays a role in cellular transformation through the Ras-MAPK pathway; and murine NIK, which interacts with upstream receptor tyrosine kinases and connects with downstream STE11-family kinases.

NEK kinases are related to NIMA, which is required for entry into mitosis in the filamentous fungus A. nidulans. Mutations in the nimA gene cause the nim (never in mitosis) G2 arrest phenotype in this fungus (Fry, A.M. and Nigg, E.A. (1995) *Current Biology* 5: 1122-1125). Several observations suggest that higher eukaryotes may have a NIMA functional counterpart(s): (1) expression of a dominant-negative form of NIMA in HeLa cells causes a G2 arrest; (2) overexpression of NIMA causes chromatin condensation, not only in A. nidulans, but also in yeast, Xenopus oocytes and HeLa cells (Lu, K.P. and Hunter, T. (1995) *Prog. Cell Cycle Res.* 1, 187-205); (3) NIMA when expressed in mammalian cells interacts with pin1, a prolyl-prolyl isomerase that functions in cell cycle regulation (Lu, K.P. *et al.* (1996) *Nature* 380, 544-547); (4) okadaic acid inhibitor studies suggests the presence of cdc2-independent mechanism to induce mitosis (Ghosh, S. *et al.*(1998) *Exp. Cell Res.* 242, 1-9) and (5) a NIMA-like kinase (fin1) exists in another eukaryote besides Aspergillus, Saccharomyces pombe (Krien, M.J.E. *et al.*(1998) *J. Cell Sci.* 111, 967-976). Eleven mammalian NIMA-like kinases have been identified – NEK1-11.

Despite the similarity of the NIMA-related kinases to NIMA over the catalytic region, the mammalian kinases are structurally different to NIMA over the extracatalytic regions. In addition several of the mammalian kinases are unable to complement the nim phenotype in Aspergillus nimA mutants.

<u>Casein Kinase 1 Group</u>

The CK1 family represents a distant branch of the protein kinase family. The hallmarks of protein kinase subdomains VIII and IX are difficult to identify. One or more forms are ubiquitously distributed in mammalian tissues and cell lines. CK1 kinases are found in cytoplasm, in nuclei, membrane-bound, and associated with the cytoskeleton. Splice variants differ in their subcellular distribution. VRK is in this group.

<u>TKL Group</u>

This group includes integrin receptor kinase (IRAK); endoribonuclease-associated kinases (IRE); Mixed lineage kinase (MLK); LIM-domain containing kinase (LIMK); MOS; PIM; Receptor interacting kinase (RIP); SR-protein specific kinase (SRPK); RAF; Serine-threonine kinase receptors (STKR).

RIP2 is a serine-threonine kinase associated with the tumor necrosis factor (TNF) receptor complex and is implicated in the activation of NF-kappa B and cell death in mammalian cells. It has recently been demonstrated that RIP2 activates the MAPK pathway (Navas, et al., *J Biol. Chem.* 1999 Nov 19;274(47): 33684-33690). RIP2 activates AP-1 and serum response element regulated expression by inducing the activation of the Elk1 transcription factor. RIP2 directly phosphorylates and activates ERK2 *in vivo* and *in vitro*. RIP2 in turn is activated through its interaction with Ras-activated Raf1. These results highlight the integrated nature of kinase signaling pathway.

<u>"Other" Group</u>

Several families cluster within a group of unrelated kinases termed "Other." Group members that define smaller, yet distinct phylogenetic branches conventional kinases include CHK1; Elongation 2 factor kinases (EIFK); Calcium-calmodulin kinase

kinases (CAMKK); IkB kinases (IKK); endoribonuclease-associated kinases (IRE);
MOS; PIM; TAK1; Testis specific kinase (TSK); tousled-related kinase (TSL);
UNC51-related kinase (UNC); WEE; mitotic kinases (BUB1, AURORA, PLK, and
NIMA/NEK); several families that are close homologues to worm (C26C2.1, YQ09,
ZC581.9, YFL033c, C24A1.3); Drosophila (SLOB), or yeast (YDOD_sp,
YGR262_sc) kinases; and others that are "unique," that is, those which do not cluster
into any obvious family. Additional families are even less well defined and first were
identified in lower eukaryotes such as yeast or worms (YNL020, YPL236, YQ09,
YWY3, SCY1, C01H6.9, C26C2.1)

The tousled (TSL) kinase was first identified in the plant Arabidopsis thaliana. TSL
encodes a serine/threonine kinase that is essential for proper flower development.
Human tousled-like kinases (Tlks) are cell-cycle-regulated enzymes, displaying
maximal activities during S phase. This regulated activity suggests that Tlk function
is linked to ongoing DNA replication (Sillje, et al., EMBO J 1999 Oct 15;18(20):
5691-5702).

BRSK Subfamily

The BRSK subfamily family of kinases includes the human BRSK1 and BRSK2,
SAD-1 from C. elegans, CG6114 from Drosophila and the HrPOPK-1 gene from the
primitive chordate Halocynthia roretzi. SAD-1 is expressed in neurons and required
for presynaptic vesicle function (Crump et al. (2001) Neuron 29: 115-29). BRSK1
and BRSK2 are selectively expressed in brain, and HrPOPK-1 is selectively expressed
in the nervous system, indicating that all members of this family have a neural
function, specifically related to synaptic vesicle function.

The NRBP family includes human kinases NRBP1 and NRBP2, as well as homologs
in C. elegans (H37N21.1) and D. melanogaster (LD28657). These kinases are most
closely related in sequence to the WNK family of kinases, and may fulfill similar
functions, including a role in hypertension.

Additionally, where BRSK2 is classsified as a member of the CAMKL family (p102), it should be further classified - i.e. "into the CAMK group, the CAMKL family and the BRSK family."

<u>Atypical Protein Kinase Group</u>

0001] There are several proteins with protein kinase activity that appear structurally unrelated to the eukaryotic protein kinases. These include; *Dictyostelium* myosin heavy chain kinase A (MHCKA), *Physarum polycephalum* actin-fragmin kinase, the human A6 PTK, human BCR, mitochondrial pyruvate dehydrogenase and branched chain fatty acid dehydrogenase kinase, and the prokaryotic "histidine" protein kinase family. The slime mold, worm, and human eEF-2 kinase homologues have all been demonstrated to have protein kinase activity, yet they bear little resemblance to conventional protein kinases except for the presence of a putative GxGxxG ATP-binding motif.

The so-called histidine kinases are abundant in prokaryotes, with more than 20 representatives in *E. coli*, and have also been identified in yeast, molds, and plants. In response to external stimuli, these kinases act as part of two-component systems to regulate DNA replication, cell division, and differentiation through phosphorylation of an aspartate in the target protein. To date, no "histidine" kinases have been identified in metazoans, although mitochondrial pyruvate dehydrogenase (PDK) and branched chain alpha-ketoacid dehydrogenase kinase (BCKD kinase), are related in sequence. PDK and BCKD kinase represent a unique family of atypical protein kinases involved in regulation of glycolysis, the citric acid cycle, and protein synthesis during protein malnutrition. Structurally they conserve only the C-terminal portion of "histidine" kinases including the G box regions. BCKD kinase phosphorylates the E1a subunit of the BCKD complex on Ser-293, proving it to be a functional protein kinase. Although no bona fide "histidine" kinase has yet been identified in humans, they do contain PDK.

Several other proteins contain protein kinase-like homology including: receptor guanylyl cyclases, diacylglycerol kinases, choline/ethanolamine kinases, and YLK1-related antibiotic resistance kinases. Each of these families contain short motifs that

were recognized by our profile searches with low scoring E-values, but a *priori* would not be expected to function as protein kinases. Instead, the similarity could simply reflect the modular nature of protein evolution and the primal role of ATP binding in diverse phosphotransfer enzymes. However, two recent papers on a bacterial homologue of the YLK1 family suggests that the aminoglycoside phosphotransferases (APHs) are structurally and functionally related to protein kinases. There are over 40 APHs identified from bacteria that are resistant to aminoglycosides such as kanamycin, gentamycin, or amikacin. The crystal structure of one well characterized APH reveals that it shares greater than 40% structural identity with the 2 lobed structure of the catalytic domain of cAMP-dependent protein kinase (PKA), including an N-terminal lobe composed of a 5-stranded antiparallel beta sheet and the core of the C-terminal lobe including several invariant segments found in all protein kinases. APHs lack the GxGxxG normally present in the loop between beta strands 1 and 2 but contain 7 of the 12 strictly conserved residues present in most protein kinases, including the HGDxxxN signature sequence in kinase subdomain VIB. Furthermore, APH also has been shown to exhibit protein-serine/threonine kinase activity, suggesting that other YLK-related molecules may indeed be functional protein kinases.

The eukaryotic lipid kinases (PI3Ks, PI4Ks, and PIPKs) also contain several short motifs similar to protein kinases, but otherwise share minimal primary sequence similarity. However, once again structural analysis of PIPKII-beta defines a conserved ATP-binding core that is strikingly similar to conventional protein kinases. Three residues are conserved among all of these enzymes including (relative to the PKA sequence) Lys-72 which binds the gamma-phosphate of ATP, Asp-166 which is part of the HRDLK motif and Asp-184 from the conserved Mg^{++} or Mn^{++} binding DFG motif. The worm genome contains 12 phosphatidylinositol kinases, including 3 PI3-kinases, 2 PI4-kinases, 3 PIP5-kinases, and 4 PI3-kinase-related kinases. The latter group has 6 mammalian members (DNA-PK, SMG1, TRRAP, FRAP/TOR, ATM, and ATR), which have been shown to participate in the maintenance of genomic integrity in response to DNA damage, and exhibit true protein kinase activity, raising the possibility that other PI-kinases may also act as protein kinases. Regardless of

-12-

whether they have true protein kinase activity, PI3-kinases are tightly linked to protein kinase signaling, as evidenced by their involvement downstream of many growth factor receptors and as upstream activators of the cell survival response mediated by the AKT protein kinase.

SUMMARY OF THE INVENTION

The present invention relates, in part, to human protein kinases and protein kinase-like enzymes identified from genomic and cDNA sequencing.

Tyrosine and serine/threonine kinases (PTK's and STK's) have been identified and their protein sequence predicted as part of the instant invention. Mammalian members of these families were identified through the use of a bioinformatics strategy. The partial or complete sequences of these kinases are presented here, together with their classification.

One aspect of the invention features an identified, isolated, enriched, or purified nucleic acid molecule encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132.

The term "identified" in reference to a nucleic acid means that a sequence was selected from a genomic, EST, or cDNA sequence database based on it being predicted to encode a portion of a previously unknown or novel protein kinase.

By "isolated," in reference to nucleic acid, is meant a polymer of 10, 15, or 18 (preferably 21, more preferably 39, most preferably 75) or more nucleotides conjugated to each other, including DNA and RNA that is isolated from a natural source or that is synthesized as the sense or complementary antisense strand. In certain embodiments of the invention, longer nucleic acids are preferred, for example those of 100, 200, 300, 400, 500, 600, 900, 1200, 1500, or more nucleotides and/or those having at least 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to a sequence selected from the group

consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66 or encoding
for amino acid selected from SEQ ID NO: 67 through 132.

The isolated nucleic acid of the present invention is unique in the sense that it is not
found in a pure or separated state in nature. Use of the term "isolated" indicates that a
naturally occurring sequence has been removed from its normal cellular (*i.e.*,
chromosomal) environment. Thus, the sequence may be in a cell-free solution or
placed in a different cellular environment. The term does not imply that the sequence
is the only nucleotide chain present, but that it is essentially free (about 90 - 95% pure
at least) of non-nucleotide material naturally associated with it, and thus is
distinguished from isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that the
specific DNA or RNA sequence constitutes a significantly higher fraction (2- to 5-
fold) of the total DNA or RNA present in the cells or solution of interest than in
normal or diseased cells or in the cells from which the sequence was taken. This
could be caused by a person by preferential reduction in the amount of other DNA or
RNA present, or by a preferential increase in the amount of the specific DNA or RNA
sequence, or by a combination of the two. However, it should be noted that enriched
does not imply that there are no other DNA or RNA sequences present, just that the
relative amount of the sequence of interest has been significantly increased. The term
"significant" is used to indicate that the level of increase is useful to the person
making such an increase, and generally means an increase relative to other nucleic
acids of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The
term also does not imply that there is no DNA or RNA from other sources. The DNA
from other sources may, for example, comprise DNA from a yeast or bacterial
genome, or a cloning vector such as pUC19. This term distinguishes from naturally
occurring events, such as viral infection, or tumor-type growths, in which the level of
one mRNA may be naturally increased relative to other species of mRNA. That is,
the term is meant to cover only those situations in which a person has intervened to
elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at least 2- to 5-fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

By a "kinase polypeptide" is meant 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids in a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132. In certain aspects, polypeptides of 75, 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more amino acids are preferred. The kinase polypeptide can be encoded by a full-length nucleic acid sequence or any portion (*e.g.*, a "fragment" as defined herein) of the full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained, including, for example, a catalytic domain, as defined herein, or a portion thereof. One of skill in the art would be able to select those catalytic domains, or portions thereof, which exhibit a kinase or kinase-like activity, *e.g.*, catalytic activity, as defined herein. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or

-15-

polypeptide which retains the functionality of the original. Such substitutions may include the replacement of an amino acid by a residue having similar physicochemical properties, such as substituting one aliphatic residue (Ile, Val, Leu or Ala) for another, or substitution between basic residues Lys and Arg, acidic residues Glu and Asp, amide residues Gln and Asn, hydroxyl residues Ser and Tyr, or aromatic residues Phe and Tyr. Further information regarding making amino acid exchanges which have only slight, if any, effects on the overall protein can be found in Bowie *et al.*, *Science*, 1990, *247*, 1306-1310, which is incorporated herein by reference in its entirety including any figures, tables, or drawings. In all cases, all permutations are intended to be covered by this disclosure.

The amino acid sequence of a kinase peptide of the invention will be substantially similar to a sequence having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, or the corresponding full-length amino acid sequence, or fragments thereof.

A sequence that is substantially similar to a sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, will preferably have at least 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the sequence.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity using standard parameters, for example Gapped BLAST or PSI-BLAST (Altschul, *et al.* (1997) *Nucleic Acids Res.* 25: 3389-3402), BLAST (Altschul, *et al.* (1990) *J. Mol. Biol.* 215: 403-410), and Smith-Waterman (Smith, *et al.* (1981) *J. Mol. Biol.* 147: 195-197). Preferably, the default settings of these programs will be employed, but those skilled in the art

-16-

recognize whether these settings need to be changed and know how to make the changes.

"Similarity" is measured by dividing the number of identical residues plus the number of conservatively substituted residues (see Bowie, *et al. Science*, 1999), 247, 1306-1310, which is incorporated herein by reference in its entirety, including any drawings, figures, or tables) by the total number of residues and gaps and multiplying the product by 100.

In preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding a kinase polypeptide comprising a nucleotide sequence that: (a) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132 or an amino acid sequence having at least about 90% identical to a sequence selected from the group consisting of SEQ ID NO: 67 through SEQ ID NO: 132; (b) is the complement of the nucleotide sequence of (a); (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring kinase polypeptide; (d) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, except that it lacks one or more, but not all, of the domains selected from the group consisting of the protein kinase, CNH, PH, phobol esters/diacylglycerol binding (C1), protein kinase C-terminal, PDZ (also known as DHR or GLGF), kinase associated domain 1, UBA/TS-N, UBA, armadillo/beta-catenin-like repeat, POLO box duplicated region, P21-Rho-binding, immunoglobulin, WIF, leucine rich repeat, SH3, MYND, EF hand, and bromodomain; (e) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, except that it lacks one or more, but not all, of the regions selected from the C-terminal region, the N-terminal region, a spacer region, and the catalytic domain; and (f) is the complement of the nucleotide sequence of (d) or (e).

The invention includes an antibody or antibody fragment having specific binding affinity to a kinase polypeptide or to a domain of said polypeptide, wherein said

-17-

polypeptide comprises an amino acid sequence selected from those set forth in SEQ ID NO: 67 through 132, a hybridoma which produces the such an antibody or antibody fragment, a kit comprising such an antibody which binds to a polypeptide of the invention a negative control antibody.

The invention includes a method for identifying a substance that modulates the activity of a kinase polypeptide comprising the steps of: (a)contacting the kinase polypeptide substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 with a test substance; (b)measuring the activity of said polypeptide; and (c)determining whether said substance modulates the activity of said polypeptide.

The invention also includes a method for identifying a substance that modulates the activity of a kinase polypeptide in a cell comprising the steps of: expressing a kinase polypeptide having a sequence substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132; adding a test substance to said cell; and monitoring a change in cell phenotype or the interaction between said polypeptide and a natural binding partner.

The invention includes a method for treating a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a kinase substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132.

The treatmnet methods of the invention include the disease or disorder is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, metabolic disorders and inflammatory disorders; and the disease or disorder selected from the group consisting of cancers of tissues; cancers of blood or hematopoietic origin; cancers of the breast, colon, lung, prostate, cervix, brain, ovaries, bladder or kidney. The treatment methods also include the disease or disorder is selected from the group consisting of disorders of the central or peripheral nervous system; migraines; pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension;

-18-

hypertension; psychotic disorders; neurological disorders and dyskinesias. Treatment methods also include disease or disorder selected from the group consisting of inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity and organ transplant rejection.

The methods of the invention contemplate use of a substance that modulates kinase activity *in vitro*, including kinase inhibitors.

The invention includes a method for detection of a kinase polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein said method comprises:

(a) contacting said sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132, said probe comprising the nucleic acid sequence, fragments thereof, or the complements of said sequences and fragments; and

(b) detecting the presence or amount of the target region: probe hybrid, as an indication of said disease or disorder.

Such a detection method includes a disease or disorder selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, metabolic disorders and inflammatory disorders; a disease or disorder selected from the group consisting of cancers of tissues; cancers of blood or hematopoietic origin; cancers of the breast, colon, lung, prostate, cervix, brain, ovary, bladder or kidney; a disease or disorder is selected from the group consisting of central or peripheral nervous system disease, migraines, pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders and dyskinesias; a disease or disorder is selected from the group consisting of inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma,

-19-

osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

The invention includes an isolated, enriched or purified nucleic acid molecule that comprises a nucleic molecule encoding a domain of a kinase polypeptide having a sequence of SEQ ID NO: 67-132.

The invention includes an isolated, enriched or purified nucleic acid molecule encoding a kinase polypeptide which comprises a nucleotide sequence that encodes a polypeptide having an amino acid sequence that has at least 90 % identity to a polypeptide set forth in SEQ ID NO: 67-132.

The invention includes an isolated, enriched or purified nucleic acid molecule according wherein the molecule comprises a nucleotide sequence substantially identical to a sequence of SEQ ID NO: 1-66.

The invention includes an isolated, enriched or purified nucleic acid molecule consisting essentially of about 10-30 contiguous nucleotide bases of a nucleic acid sequence that encodes a polypeptide selected from the group consisting of SEQ ID NO: 67 through 132. The invention also includes an isolated, enriched or purified nucleic acid molecule of about 10-30 contiguous nucleotide bases of a nucleic acid sequence that encodes a polypeptide selected from the group consisting of SEQ ID NO: 67 through 132, consisting essentially of about 10-30 contiguous nucleotide bases of a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 66.

The term "complement" refers to two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if all of the nucleotides of the first sequence are complementary to all of the nucleotides of the second sequence.

Various low or high stringency hybridization conditions may be used depending upon
the specificity and selectivity desired. These conditions are well known to those
skilled in the art. Under stringent hybridization conditions only highly
complementary nucleic acid sequences hybridize. Preferably, such conditions prevent
hybridization of nucleic acids having more than 1 or 2 mismatches out of 20
contiguous nucleotides, more preferably, such conditions prevent hybridization of
nucleic acids having more than 1 or 2 mismatches out of 50 contiguous nucleotides,
most preferably, such conditions prevent hybridization of nucleic acids having more
than 1 or 2 mismatches out of 100 contiguous nucleotides. In some instances, the
conditions may prevent hybridization of nucleic acids having more than 5 mismatches
in the full-length sequence.

By stringent hybridization assay conditions is meant hybridization assay conditions at
least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM
NaH2PO4, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X
Denhardt's solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C;
and washing with 0.2X SSC, 0.1% SDS at 45 °C. Under some of the most stringent
hybridization assay conditions, the second wash can be done with 0.1X SSC at a
temperature up to 70 °C (Berger *et al.* (1987) Guide to Molecular Cloning Techniques
pg 421, hereby incorporated by reference herein in its entirety including any figures,
tables, or drawings.). However, other applications may require the use of conditions
falling between these sets of conditions. Methods of determining the conditions
required to achieve desired hybridizations are well known to those with ordinary skill
in the art, and are based on several factors, including but not limited to, the sequences
to be hybridized and the samples to be tested. Washing conditions of lower
stringency frequently utilize a lower temperature during the washing steps, such as 65
°C, 60 °C, 55 °C, 50 °C, or 42 °C.

The term "domain" refers to a region of a polypeptide whose sequence or structure is
conserved between several homologs of the polypoeptide and which serves a
particular function. Many domains may be identified by searching the Pfam database
of domain models (http: //pfam.wustl.edu) which provides coordinates on the

polypeptide delimiting the start and end of the domain, as well as a score giving the likelihood that the domain is present in the polypeptide. Other domains may be identified by specialized programs, such as the COILS program to detect colied-coil regions (http: //www.ch.embnet.org/software/COILS_form.html), the SignalP program to detect signal peptides (http: //www.ebs.dtu.dk/services/TMIIMM), by visual inspection of the amino acid sequence (e.g., determination of cysteine-rich or proline-rich domains), or by Smith-Waterman alignment shows a high level of sequence similarity in the region containing the domain, it may be concluded that the domain is present in both proteins within that region. which serves a particular function.

Domains of signal transduction proteins can serve functions including, but not limited to, binding molecules that localize the signal transduction molecule to different regions of the cell, binding other signaling molecules directly responsible for propagating a particular cellular signal or binding molecules that influence the function of the protein. Some domains can be expressed separately from the rest of the protein and function by themselves

The term "N-terminal region" refers to the extracatalytic region located between the initiator methionine and the catalytic domain of the protein kinase. Depending on its length, the N-terminal region may or may not play a regulatory role in kinase function. An example of a protein kinase whose N-terminal domain has been shown to play a regulatory role is PAK6 or PAK5, which contains a CRIB motif used for Cdc42 and rac binding (Burbelo, P.D. *et al.* (1995) *J. Biol. Chem.* 270, 29071-29074). Such an N-terminal region is also termed a N-terminal functional domain or N-terminal domain.

The term "catalytic domain" or protein kinase domain refers to a region of the protein kinase that is typically 25-300 amino acids long and is responsible for carrying out the phosphate transfer reaction from a high-energy phosphate donor molecule such as ATP or GTP to itself (autophosphorylation) or to other proteins (exogenous phosphorylation). The catalytic domain of protein kinases is made up of 12 subdomains that contain highly conserved amino acid residues, and are responsible

-22-

for proper polypeptide folding and for catalysis. The catalytic dmoain can be defined with reference to the parameters described in a "Pfam" database: http: //pfam.wustl.edu. In particular, it can be defined with reference to a HMMer search of the Pfam database. In the N-terminal extremity of the catalytic domain there is a glycine rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. In the central part of the catalytic domain there is a conserved aspartic acid residue which is important for the catalytic activity of the enzyme. See Accession number PF00069 of http: //pfam.wustl.edu.

The term "catalytic activity," as used herein, defines the rate at which a kinase catalytic domain phosphorylates a substrate. Catalytic activity can be measured, for example, by determining the amount of a substrate converted to a phosphorylated product as a function of time. Catalytic activity can be measured by methods of the invention by determining the concentration of a phosphorylated substrate after a fixed period of time. Phosphorylation of a substrate occurs at the active site of a protein kinase. The active site is normally a cavity in which the substrate binds to the protein kinase and is phosphorylated.

The term "substrate" as used herein refers to a molecule phosphorylated by a kinase of the invention. Kinases phosphorylate substrates on serine/threonine or tyrosine amino acids. The molecule may be another protein or a polypeptide.

The term "C-terminal region" refers to the region located between the catalytic domain or the last (located closest to the C-terminus) functional domain and the carboxy-terminal amino acid residue of the protein kinase. See Accession number PF00433 of http: //pfam.wustl.edu. Depending on its length and amino acid composition, the C-terminal region may or may not play a regulatory role in kinase function. An example of a protein kinase whose C-terminal region may play a regulatory role is PAK3 which contains a heterotrimeric G_b subunit-binding site near its C-terminus (Leeuw, T. *et al.* (1998) *Nature*, 391, 191-195). Such a C-terminal region is also termed a C-terminal functional domain or C-terminal domain.

-23-

By "functional" domain is meant any region of the polypeptide that may play a regulatory or catalytic role as predicted from amino acid sequence homology to other proteins or by the presence of amino acid sequences that may give rise to specific structural conformations.

The "CNH domain" is the citron homology domain, and is often found after cysteine rich and pleckstrin homology (PH) domains at the C-terminal end of the proteins [MEDLINE: 99321922]. It acts as a regulatory domain and could be involved in macromolecular interactions [MEDLINE: 99321922], [MEDLINE: 97280817]. See Accession number PF00780 of http: //pfam.wustl.edu.

The "PH domain" is the 'pleckstrin homology' (PH) domain and is a domain of about 100 residues that occurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton [MEDLINE: 93272305], [MEDLINE: 93268380], [MEDLINE: 94054654], [MEDLINE: 95076505], [MEDLINE: 95157628], [MEDLINE: 95197706], [MEDLINE: 96082954]. See Accession number PF00169 of http: //pfam.wustl.edu.

The "Phorbol esters/diacylglycerol binding domain" is also known as the Protein kinase C conserved region 1 (C1) domain. The N-terminal region of PKC, known as C1, has been shown [MEDLINE: 89296905] to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. See Accession number PF00130 of http: //pfam.wustl.edu.

The "PDZ domain" is also known as the DHR or GLGF domain. PDZ domains are found in diverse signaling proteins and may function in targeting signalling molecules to sub-membranous sites [MEDLINE: 97348826]. See Accession number PF00595 of http: //pfam.wustl.edu.

The "kinase associated domain 1" (KA1) domain is found in the C-terminal extremity of various serine/threonine-protein kinases from fungi, plants and animals. See Accession number PF02149 of http: //pfam.wustl.edu.

The UBA/TS-N domain is composed of three alpha helices. This family includes the previously defined UBA and TS-N domains. The UBA-domain (ubiquitin associated domain) is a sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway. The structure of the UBA domain consists of a compact three helix bundle. This domain is found at the N terminus of EF-TS hence the name TS-N. The structure of EF-TS is known and this domain is implicated in its interaction with EF-TU. The domain has been found in non EF-TS proteins such as alpha-NAC P70670 and MJ0280 Q57728 [1]. See Accession number PF00627 of http: //pfam.wustl.edu.

The "UBA domain" The UBA-domain (ubiquitin associated domain) is a novel sequence motif found in several proteins having connections to ubiquitin and the ubiquitination pathway [MEDLINE: 97025177]. The UBA domain is probably a non-covalent ubiquitin binding domain consisting of a compact three helix bundle [MEDLINE: 99061330]. See Accession number PF00627 of http: //pfam.wustl.edu.

The "armadillo/beta-catenin-like repeat" is an approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity gene armadillo. Similar repeats were later found in the mammalian armadillo homolog beta-catenin, the junctional plaque protein plakoglobin, the adenomatous polyposis coli (APC) tumor suppressor protein, and a number of other proteins [MEDLINE: 94170379]. The 3 dimensional fold of an armadillo repeat is known from the crystal structure of beta-catenin [MEDLINE: 98449700]. There, the 12 repeats form a superhelix of alpha-helices, with three helices per unit. The cylindrical structure features a positively charged grove which presumably interacts with the acidic surfaces of the known interaction partners of beta-catenin. See Accession number PF00514 of http: //pfam.wustl.edu.

-25-

The "POLO box duplicated region" (POLO_box) is described as follows. A subgroup of serine/threonine protein kinases (IPR002290) playing multiple roles during cell cycle, especially in M phase progression and cytokinesis, contain a duplicated domain in their C terminal part, the polo box [MEDLINE: 99116035]. The domain is named after its founding member encoded by the polo gene of Drosophila [MEDLINE: 92084090]. This domain of around 70 amino acids has been found in species ranging from yeast to mammals. Point mutations in the Polo box of the budding yeast Cdc5 protein abolish the ability of overexpressed Cdc5 to interact with the spindle poles and to organize cytokinetic structures [MEDLINE: 20063188]. See Accession number PF00659 of http: //pfam.wustl.edu.

The "P21-Rho-binding domain" is one of a group of small domains that bind Cdc42p-and/or Rho-like small GTPases. These are also known as the Cdc42/Rac interactive binding (CRIB). See Accession number PF00786 of http: //pfam.wustl.edu.

The "immunoglobulin domain" is a domain that is under the umbrella of the immunoglobulin superfamily. Examples of the superfamily include antibodies, the giant muscle kinase titin and receptor tyrosine kinases. Immunoglobulin-like domains may be involved in protein-protein and protein-ligand interactions. The Pfam alignments do not include the first and last strand of the immunoglobulin-like domain. See Accession number PF00047 of http: //pfam.wustl.edu.

The "WIF domain" is found in the RYK tyrosine kinase receptors and WIF the Wnt-inhibitory-factor. The domain is extracellular and and contains two conserved cysteines that may form a disulphide bridge. This domain is Wnt binding in WIF, and it has been suggested that RYK may also bind to Wnt [MEDLINE: 20105592]. See Accession number PF02019 of http: //pfam.wustl.edu.

The "leucine rich repeat" - Leucine-rich repeats (LRRs) are relatively short motifs (22-28 residues in length) found in a variety of cytoplasmic, membrane and extracellular proteins [MEDLINE: 91099665]. Although these proteins are associated with widely different functions, a common property involves protein-protein interaction. Other functions of LRR-containing proteins include, for example, binding

-26-

to enzymes [MEDLINE: 90094386] and vascular repair [MEDLINE: 89367331].
See Accession number PF00560 of http: //pfam.wustl.edu.

The "SH3 domain" SH3 (src Homology-3) domains are small protein modules
containing approximately 50 amino acid residues [PUB00001025]. They are found in
a variety of of proteins with enzymatic activity. The SH3 domain has a characteristic
fold which consists of five or six beta-strands arranged as two tightly packed anti-
parallel beta sheets. The linker regions may contain short helices [PUB00001083].
See Accession number PF00018 of http: //pfam.wustl.edu.

The "MYND finger" is a domain found in some suppressors of cell cycle entry
[MEDLINE: 96203118], [MEDLINE: 98079069]. The MYND zinc finger (ZnF)
domain is one of two domains in AML/ETO fusion protein required for repression of
basal transcription from the multidrug resistance 1 (MDR-1) promoter. The other
domain is a hydrophobic heptad repeat (HHR) motif [MEDLINE: 98252948]. The
AML-1/ETO fusion protein is created by the (8;21) translocation, the second most
frequent chromosomal abnormality associated with acute myeloid leukemia. In the
fusion protein the AML-1 runt homology domain, which is responsible for DNA
binding and CBF beta interaction, is linked to ETO, a gene of unknown function
[MEDLINE: 96068903]. See Accession number PF01753 of http: //pfam.wustl.edu.

The "EF hand" domain is described as follows: many calcium-binding proteins
belong to the same evolutionary family and share a type of calcium-binding domain
known as the EF-hand. This type of domain consists of a twelve residue loop flanked
on both side by a twelve residue alpha-helical domain. In an EF-hand loop the
calcium ion is coordinated in a pentagonal bipyramidal configuration. The six
residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are
denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides
two oxygens for liganding Ca (bidentate ligand). See Accession number PF00036 of
http: //pfam.wustl.edu.

A "bromodomain" is a 110 amino acid long domain, found in many chromatin
associated proteins. Bromodomains can interact specifically with acetylated lysine.

[MEDLINE: 97318593] Bromodomains are found in a variety of mammalian, invertebrate and yeast DNA-binding proteins [MEDLINE: 92285152].. The bromodomain may occur as a single copy, or in duplicate. The bromodomain may be involved in protein-protein interactions and may play a role in assembly or activity of multi-component complexes involved in transcriptional activation [MEDLINE: 96022440]. See Accession number PF00439 of http: //pfam.wustl.edu.

The term "coiled-coil structure region" as used herein, refers to a polypeptide sequence that has a high probability of adopting a coiled-coil structure as predicted by computer algorithms such as COILS (Lupas, A. (1996) *Meth. Enzymology* 266: 513-525). Coiled-coils are formed by two or three amphipathic α-helices in parallel. Coiled-coils can bind to coiled-coil domains of other polypeptides resulting in homo- or heterodimers (Lupas, A. (1991) *Science* 252: 1162-1164). Coiled-coil-dependent oligomerization has been shown to be necessary for protein function including catalytic activity of serine/threonine kinases (Roe, J. *et al.* (1997) *J. Biol. Chem.* 272: 5838-5845).

The term "proline-rich region" as used herein, refers to a region of a protein kinase whose proline content over a given amino acid length is higher than the average content of this amino acid found in proteins(*i.e.*, >10%). Proline-rich regions are easily discernable by visual inspection of amino acid sequences and quantitated by standard computer sequence analysis programs such as the DNAStar program EditSeq. Proline-rich regions have been demonstrated to participate in regulatory protein -protein interactions. Among these interactions, those that are most relevant to this invention involve the "PxxP" proline rich motif found in certain protein kinases (*i.e.*, human PAK1) and the SH3 domain of the adaptor molecule Nck (Galisteo, M.L. *et al.* (1996) *J. Biol. Chem.* 271: 20997-21000). Other regulatory interactions involving "PxxP" proline-rich motifs include the WW domain (Sudol, M. (1996) *Prog. Biochys. Mol. Bio.* 65: 113-132).

The term "spacer region" as used herein, refers to a region of the protein kinase located between predicted functional domains. The spacer region has little conservation when compared with any any amino acid sequence in the database, and

-28-

can be identified by using a Smith-Waterman alignment of the protein sequence against the non-redundant protein of Pfam database to define the C- and N-terminal boundaries of the flanking functional domains. Spacer regions may or may not play a fundamental role in protein kinase function. Precedence for the regulatory role of spacer regions in kinase function is provided by the role of the *src* kinase spacer in inter-domain interactions (Xu, W. *et al.* (1997) *Nature* 385: 595-602).

The term "insert" as used herein refers to a portion of a protein kinase that is absent from a close homolog. Inserts may or may not by the product alternative splicing of exons. Inserts can be identified by using a Smith-Waterman sequence alignment of the protein sequence against the non-redundant protein database, or by means of a multiple sequence alignment of homologous sequences using the DNAStar program Megalign. Inserts may play a functional role by presenting a new interface for protein-protein interactions, or by interfering with such interactions.

The term "signal transduction pathway" refers to the molecules that propagate an extracellular signal through the cell membrane to become an intracellular signal. This signal can then stimulate a cellular response. The polypeptide molecules involved in signal transduction processes are typically receptor and non-receptor protein kinases, receptor and non-receptor protein phosphatases, polypeptides containing *SRC* homology 2 and 3 domains, phosphotyrosine binding proteins (*SRC* homology 2 (SH2) and phosphotyrosine binding (PTB and PH) domain containing proteins), proline-rich binding proteins (SH3 domain containing proteins), GTPases, phosphodiesterases, phospholipases, prolyl isomerases, proteases, Ca2+ binding proteins, cAMP binding proteins, guanyl cyclases, adenylyl cyclases, NO generating proteins, nucleotide exchange factors, and transcription factors.

In other preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding kinase polypeptides, further comprising a vector or promoter effective to initiate transcription in a host cell. The nucleic acid may encode a polypeptide of SEQ ID NO: 67-132 and a vector or promoter effective to initiate transcription in a host cell. The invention includes such nucleic acid molecules that are isolated, enriched, or purified from a mammal and in a preferred embodiment, the

-29-

mammal is a human. The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence selected from the group consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66, or a functional derivative thereof and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence complementary to an RNA sequence encoding a kinase polypeptide and a transcriptional termination region functional in a cell. Specific vectors and host cell combinations are discussed herein.

The term "vector" relates to a single or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a kinase can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together.

The term "transfecting" defines a number of methods to insert a nucleic acid vector or other nucleic acid molecules into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt, an electric field, detergent, or DMSO to render the outer membrane or wall of the cells permeable to nucleic acid molecules of interest or use of various viral transduction strategies.

The term "promoter" as used herein, refers to nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to persons skilled in the art for different organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally

-30-

WO 2004/006838

PCT/US2003/021730

include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

In preferred embodiments, the isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence selected from the group consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66, which encodes an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132 , a functional derivative thereof, or at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, the catalytic region of SEQ ID NO: 67-132 or catalytic domains, functional domains, or spacer regions of SEQ ID NO: 67 through 132. The nucleic acid may be isolated from a natural source by cDNA cloning or by subtractive hybridization. The natural source may be mammalian, preferably human, preferably blood, semen or tissue, and the nucleic acid may be synthesized by the triester method or by using an automated DNA synthesizer.

The term "mammal" refers preferably to such organisms as mice, rats, rabbits, guinea pigs, sheep, and goats, more preferably to cats, dogs, monkeys, and apes, and most preferably to humans.

In yet other preferred embodiments, the nucleic acid is a conserved or unique region, for example those useful for: the design of hybridization probes to facilitate identification and cloning of additional polypeptides, the design of PCR probes to facilitate cloning of additional polypeptides, obtaining antibodies to polypeptide regions, and designing antisense oligonucleotides.

By "conserved nucleic acid regions," are meant regions present on two or more nucleic acids encoding a kinase polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acid encoding kinase polypeptides are provided in Wahl *et al. Meth. Enzym.* 152: 399-407 (1987) and in Wahl *et al. Meth. Enzym.* 152: 415-423 (1987), which are hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables. Preferably,

-31-

BNSDOCID: <WO____2004006838A2_I_>

conserved regions differ by no more than 5 out of 20 nucleotides, even more preferably 2 out of 20 nucleotides or most preferably 1 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a nucleic acid coding for a kinase polypeptide that is not present in a sequence coding for any other naturally occurring polypeptide. Such regions preferably encode 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids, for example, an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132. In particular, a unique nucleic acid region is preferably of mammalian origin.

Another aspect of the invention features a nucleic acid probe for the detection of nucleic acid encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, catalytic domains, functional domains, or spacer regions of SEQ ID NO: 67 through 132, in a sample. The nucleic acid probe contains a nucleotide base sequence that will hybridize to the sequence selected from the group consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66, a sequence encoding catalytic domains, functional domains, or spacer regions of SEQ ID NO: 67 through 132, or a functional derivative thereof.

In preferred embodiments, the nucleic acid probe hybridizes to nucleic acid encoding at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids, wherein the nucleic acid sequence is selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 66, or a functional derivative thereof.

Methods for using the probes include detecting the presence or amount of kinase RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to kinase RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a kinase polypeptide may be used in the identification of the sequence of the nucleic acid detected (Nelson *et al.*, *in Nonisotopic DNA Probe Techniques*, Academic Press, San Diego, Kricka, ed., p. 275,

-32-

1992, hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

Methods for using the probes also include using these probes to find, for example, the full-length clone of each of the predicted kinases by techniques known to one skilled in the art. These clones will be useful for screening for small molecule compounds that inhibit the catalytic activity of the encoded kinase with potential utility in treating cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically disorders including cancers of tissues or blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervix, skin, brain, ovary, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, multiple sclerosis, and amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, hypertension, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, bone disorder, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

In another aspect, the invention describes a recombinant cell or tissue comprising a nucleic acid molecule encoding a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132.

-33-

In such cells, the nucleic acid may be under the control of the genomic regulatory elements, or may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not normally coupled *in vivo* transcriptionally to the coding sequence for the kinase polypeptides.

The polypeptide is preferably a fragment of the protein encoded by an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132. By "fragment," is meant an amino acid sequence present in a kinase polypeptide. Preferably, such a sequence comprises at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino acids of a sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132.

In another aspect, the invention features an isolated, enriched, or purified kinase polypeptide having the amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132.

By "isolated" in reference to a polypeptide is meant a polymer of 6 (preferably 12, more preferably 18, or 21, most preferably 25, 32, 40, or 50) or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are synthesized. In certain aspects longer polypeptides are preferred, such as those comprising 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more contiguous amino acids, including an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 ; other longer polypeptides also preferred are those having sequence that is substantially similar to a sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through SEQ ID NO: 132(which preferably has at least 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the sequence).

The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only

-34-

amino acid chain present, but that it is essentially free (about 90 - 95% pure at least) of non-amino acid-based material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence constitutes a significantly higher fraction (2- to 5-fold) of the total amino acid sequences present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acid sequences present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significantly" here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acid sequences of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no amino acid sequence from other sources. The other source of amino acid sequences may, for example, comprise amino acid sequence encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened to increase the proportion of the desired amino acid sequence.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment. Compared to the natural level this level should be at least 2-to 5-fold greater (e.g., in terms of mg/mL). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure.

-35-

In preferred embodiments, the kinase polypeptide is a fragment of the protein encoded by an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 . Preferably, the kinase polypeptide contains at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino acids of a sequence selected from the group consisting of those set forth in SEQ ID NO: 3 and 4 , or a functional derivative thereof.

In preferred embodiments, the kinase polypeptide comprises an amino acid sequence having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 ; and (b) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 , except that it lacks one or more of the domains selected from the group consisting of the catalytic domain, the C-terminal region, the N-terminal region, and the spacer region.

The polypeptide can be isolated from a natural source by methods well-known in the art. The natural source may be mammalian, preferably human, preferably blood, semen or tissue, and the polypeptide may be synthesized using an automated polypeptide synthesizer.

In some embodiments the invention includes a recombinant kinase polypeptide having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 . By "recombinant kinase polypeptide" is meant a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (e.g., present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

The polypeptides to be expressed in host cells may also be fusion proteins which include regions from heterologous proteins. Such regions may be included to allow, e.g., secretion, improved stability, or facilitated purification of the polypeptide. For example, a sequence encoding an appropriate signal peptide can be incorporated into expression vectors. A DNA sequence for a signal peptide (secretory leader) may be

-36-

fused in-frame to the polynucleotide sequence so that the polypeptide is translated as a fusion protein comprising the signal peptide. A signal peptide that is functional in the intended host cell promotes extracellular secretion of the polypeptide. Preferably, the signal sequence will be cleaved from the polypeptide upon secretion of the polypeptide from the cell. Thus, preferred fusion proteins can be produced in which the N-terminus of a kinase polypeptide is fused to a carrier peptide.

In one embodiment, the polypeptide comprises a fusion protein which includes a heterologous region used to facilitate purification of the polypeptide. Many of the available peptides used for such a function allow selective binding of the fusion protein to a binding partner. A preferred binding partner includes one or more of the IgG binding domains of protein A are easily purified to homogeneity by affinity chromatography on, for example, IgG-coupled Sepharose. Alternatively, many vectors have the advantage of carrying a stretch of histidine residues that can be expressed at the N-terminal or C-terminal end of the target protein, and thus the protein of interest can be recovered by metal chelation chromatography. A nucleotide sequence encoding a recognition site for a proteolytic enzyme such as enterokinase, factor X procollagenase or thrombine may immediately precede the sequence for a kinase polypeptide to permit cleavage of the fusion protein to obtain the mature kinase polypeptide. Additional examples of fusion-protein binding partners include, but are not limited to, the yeast I-factor, the honeybee melatin leader in sf9 insect cells, 6-His tag, thioredoxin tag, hemaglutinin tag, GST tag, and OmpA signal sequence tag. As will be understood by one of skill in the art, the binding partner which recognizes and binds to the peptide may be any ion, molecule or compound including metal ions (*e.g.*, metal affinity columns), antibodies, or fragments thereof, and any protein or peptide which binds the peptide, such as the FLAG tag.

In another aspect, the invention features an antibody (*e.g.*, a monoclonal or polyclonal antibody) having specific binding affinity to a kinase polypeptide or a kinase polypeptide domain or fragment where the polypeptide is selected from the group having a sequence at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to an amino acid sequence set forth in SEQ ID NO: 67

through 132. By "specific binding affinity" is meant that the antibody binds to the target kinase polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody fragments are polypeptides that contain regions that can bind other polypeptides. Antibodies can be used to identify an endogenous source of kinase polypeptides, to monitor cell cycle regulation, and for immuno-localization of kinase polypeptides within the cell.

The term "polyclonal" refers to antibodies that are heterogenous populations of antibody molecules derived from the sera of animals immunized with an antigen or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.

"Monoclonal antibodies" are substantially homogenous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained by methods known to those skilled in the art (Kohler *et al., Nature* 256: 495-497, 1975, and U.S. Patent No. 4,376,110, both of which are hereby incorporated by reference herein in their entirety including any figures, tables, or drawings).

An antibody of the present invention includes "humanized" monoclonal and polyclonal antibodies. Humanized antibodies are recombinant proteins in which non-human (typically murine) complementarity determining regions of an antibody have been transferred from heavy and light variable chains of the non-human (*e.g.* murine) immunoglobulin into a human variable domain, followed by the replacement of some human residues in the framework regions of their murine counterparts. Humanized antibodies in accordance with this invention are suitable for use in therapeutic methods. General techniques for cloning murine immunoglobulin variable domains are described, for example, by the publication of Orlandi *et al., Proc. Nat'l Acad. Sci. USA 86*: 3833 (1989). Techniques for producing humanized monoclonal antibodies are described, for example, by Jones *et al., Nature 321*: 522 (1986), Riechmann *et al.,*

-38-

Nature 332: 323 (1988), Verhoeyen *et al., Science 239*: 1534 (1988), Carter *et al., Proc. Nat'l Acad. Sci. USA 89*: 4285 (1992), Sandhu, *Crit. Rev. Biotech. 12*: 437 (1992), and Singer *et al., J. Immun. 150*: 2844 (1993).

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

An antibody fragment of the present invention includes a "single-chain antibody," a phrase used in this description to denote a linear polypeptide that binds antigen with specificity and that comprises variable or hypervariable regions from the heavy and light chains of an antibody. Such single chain antibodies can be produced by conventional methodology. The Vh and Vl regions of the Fv fragment can be covalently joined and stabilized by the insertion of a disulfide bond. See Glockshuber, *et al., Biochemistry* 1362 (1990). Alternatively, the Vh and Vl regions can be joined by the insertion of a peptide linker. A gene encoding the Vh, Vl and peptide linker sequences can be constructed and expressed using a recombinant expression vector. See Colcher, *et al., J. Nat'l Cancer Inst.* 82: 1191 (1990). Amino acid sequences comprising hypervariable regions from the Vh and Vl antibody chains can also be constructed using disulfide bonds or peptide linkers.

Antibodies or antibody fragments having specific binding affinity to a polypeptide of the invention may be used in methods for detecting the presence and/or amount of kinase polypeptide in a sample by probing the sample with the antibody under conditions suitable for kinase antibody immunocomplex formation and detecting the presence and/or amount of the antibody conjugated to the kinase polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for the kinase as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

An antibody or antibody fragment with specific binding affinity to a kinase polypeptide of the invention can be isolated, enriched, or purified from a prokaryotic

or eukaryotic organism. Routine methods known to those skilled in the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above. The antibody may be directly labelled with a fluorescent or radioactive label.

Antibodies having specific binding affinity to a kinase polypeptide of the invention may be used in methods for detecting the presence and/or amount of kinase polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the kinase polypeptide. Diagnostic kits for performing such methods may be constructed to include a first container containing the antibody and a second container having a conjugate of a binding partner of the antibody and a label, such as, for example, a radioisotope or fluorescent label. The diagnostic kit may also include notification of an FDA approved use and instructions therefor. Antibodies may identify phosphorylated regions of a kinase polypeptide when a protein is phosphorylated.

In another aspect, the invention features a hybridoma which produces an antibody having specific binding affinity to a kinase polypeptide or a kinase polypeptide domain, where the polypeptide is selected from the group having an amino acid sequence set forth in SEQ ID NO: 67 through 132. By hybridoma is meant an immortalized cell line that is capable of secreting an antibody, for example an antibody to a kinase of the invention. In preferred embodiments, the antibody to the kinase comprises a sequence of amino acids that is able to specifically bind a kinase polypeptide of the invention.

In another aspect, the present invention is also directed to kits comprising antibodies that bind to a polypeptide encoded by any of the nucleic acid molecules described above, and a negative control antibody.

-40-

[0002] The term "negative control antibody" refers to an antibody derived from similar source as the antibody having specific binding affinity, but where it displays no binding affinity to a polypeptide of the invention.

In another aspect, the invention features a kinase polypeptide binding agent able to bind to a kinase polypeptide selected from the group having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 . The binding agent is preferably a purified antibody that recognizes an epitope present on a kinase polypeptide of the invention. Other binding agents include molecules that bind to kinase polypeptides and analogous molecules that bind to a kinase polypeptide. Such binding agents may be identified by using assays that measure kinase binding partner activity, such as those that measure PDGFR activity.

The invention also features a method for screening for human cells containing a kinase polypeptide of the invention or an equivalent sequence. The method involves identifying the novel polypeptide in human cells using techniques that are routine and standard in the art, such as those described herein for identifying the kinases of the invention (e.g., cloning, Southern or Northern blot analysis, in situ hybridization, PCR amplification, etc.).

In another aspect, the invention features methods for identifying a substance that modulates kinase activity comprising the steps of: (a) contacting a kinase polypeptide selected from the group having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 with a test substance; (b) measuring the activity of said polypeptide; and (c) determining whether said substance modulates the activity of said polypeptide. The skilled artisan will appreciate that the kinase polypeptides of the invention, including, for example, a portion of a full-length sequence such as a catalytic domain or a portion thereof, are useful for the identification of a substance which modulates kinase activity. Those kinase polypeptides having a functional activity (e.g., catalytic activity as defined herein) are useful for identifying a substance that modulates kinase activity.

The term "modulates" refers to the ability of a compound to alter the function of a kinase of the invention. A modulator preferably activates or inhibits the activity of a kinase of the invention depending on the concentration of the compound (modulator) exposed to the kinase.

The term "modulates" also refers to altering the function of kinases of the invention by increasing or decreasing the probability that a complex forms between the kinase and a natural binding partner. A modulator preferably increases the probability that such a complex forms between the kinase and the natural binding partner, more preferably increases or decreases the probability that a complex forms between the kinase and the natural binding partner depending on the concentration of the compound (modulator) exposed to the kinase, and most preferably decreases the probability that a complex forms between the kinase and the natural binding partner.

The term "activates" refers to increasing the cellular activity of the kinase. The term inhibit refers to decreasing the cellular activity of the kinase. Kinase activity is the phosphorylation of a substrate or the binding with a natural binding partner.

The term "complex" refers to an assembly of at least two molecules bound to one another. Signal transduction complexes often contain at least two protein molecules bound to one another. For instance, a tyrosine receptor protein kinase, GRB2, SOS, RAF, and RAS assemble to form a signal transduction complex in response to a mitogenic ligand.

The term "natural binding partner" refers to polypeptides, lipids, small molecules, or nucleic acids that bind to kinases in cells. A change in the interaction between a kinase and a natural binding partner can manifest itself as an increased or decreased probability that the interaction forms, or an increased or decreased concentration of kinase/natural binding partner complex.

The term "contacting" as used herein refers to mixing a solution comprising the test compound with a liquid medium bathing the cells of the methods. The solution comprising the compound may also comprise another component, such as dimethyl sulfoxide (DMSO), which facilitates the uptake of the test compound or compounds

-42-

into the cells of the methods. The solution comprising the test compound may be
added to the medium bathing the cells by utilizing a delivery apparatus, such as a
pipette-based device or syringe-based device.

In another aspect, the invention features methods for identifying a substance that
modulates kinase activity in a cell comprising the steps of: (a) expressing a kinase
polypeptide in a cell, wherein said polypeptide is selected from the group having an
amino acid sequence selected from the group consisting of those set forth in SEQ ID
NO: 67 through 132 ; (b) adding a test substance to said cell; and (c) monitoring a
change in kinase activity or a change in cell phenotype or the interaction between said
polypeptide and a natural binding partner. The skilled artisan will appreciate that the
kinase polypeptides of the invention, including, for example, a portion of a full-length
sequence such as a catalytic domain or a portion thereof, and are useful for the
identification of a substance which modulates kinase activity. Those kinase
polypeptides having a functional activity (*e.g.,* catalytic activity as defined herein) are
useful for identifying a substance that modulates kinase activity.

The term "expressing" as used herein refers to the production of kinases of the
invention from a nucleic acid vector containing kinase genes within a cell. The
nucleic acid vector is transfected into cells using well known techniques in the art as
described herein.

Another aspect of the instant invention is directed to methods of identifying
compounds that bind to kinase polypeptides of the present invention, comprising
contacting the kinase polypeptides with a compound, and determining whether the
compound binds the kinase polypeptides. Binding can be determined by binding
assays which are well known to the skilled artisan, including, but not limited to, gel-
shift assays, Western blots, radiolabeled competition assay, phage-based expression
cloning, co-fractionation by chromatography, co-precipitation, cross linking,
interaction trap/two-hybrid analysis, southwestern analysis, ELISA, and the like,
which are described in, for example, *Current Protocols in Molecular Biology*, 1999,
John Wiley & Sons, NY, which is incorporated herein by reference in its entirety.

-43-

The compounds to be screened include, but are not limited to, compounds of extracellular, intracellular, biological or chemical origin.

The methods of the invention also embrace compounds that are attached to a label, such as a radiolabel (e.g., ^{125}I, ^{35}S, ^{32}P, ^{33}P, ^{3}H), a fluorescence label, a chemiluminescent label, an enzymic label and an immunogenic label. The kinase polypeptides employed in such a test may either be free in solution, attached to a solid support, borne on a cell surface, located intracellularly or associated with a portion of a cell. One skilled in the art can, for example, measure the formation of complexes between a kinase polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a kinase polypeptide and its substrate caused by the compound being tested.

Other assays can be used to examine enzymatic activity including, but not limited to, photometric, radiometric, HPLC, electrochemical, and the like, which are described in, for example, _Enzyme Assays: A Practical Approach_, eds. R. Eisenthal and M. J. Danson, 1992, Oxford University Press, which is incorporated herein by reference in its entirety.

Another aspect of the present invention is directed to methods of identifying compounds which modulate (i.e., increase or decrease) activity of a kinase polypeptide comprising contacting the kinase polypeptide with a compound, and determining whether the compound modifies activity of the kinase polypeptide. As described herein, the kinase polypeptides of the invention include a portion of a full-length sequence, such as a catalytic domain, as defined herein. In some instances, the kinase polypeptides of the invention comprise less than the entire catalytic domain, yet exhibit kinase or kinase-like activity. These compounds are also referred to as "modulators of protein kinases." The activity in the presence of the test compound is compared to the activity in the absence of the test compound. Where the activity of a sample containing the test compound is higher than the activity in a sample lacking the test compound, the compound will have increased the activity. Similarly, where the activity of a sample containing the test compound is lower than the activity in the sample lacking the test compound, the compound will have inhibited the activity.

The present invention is particularly useful for screening compounds by using a kinase polypeptide in any of a variety of drug screening techniques. The compounds to be screened include, but are not limited to, extracellular, intracellular, biological or chemical origin. The kinase polypeptide employed in such a test may be in any form, preferably, free in solution, attached to a solid support, borne on a cell surface or located intracellularly. One skilled in the art can, for example, measure the formation of complexes between a kinase polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a kinase polypeptide and its substrate caused by the compound being tested.

The activity of kinase polypeptides of the invention can be determined by, for example, examining the ability to bind or be activated by chemically synthesised peptide ligands. Alternatively, the activity of the kinase polypeptides can be assayed by examining their ability to bind metal ions such as calcium, hormones, chemokines, neuropeptides, neurotransmitters, nucleotides, lipids, and odorants. Thus, modulators of the kinase polypeptide's activity may alter a kinase function, such as a binding property of a kinase or an activity such as signal transduction or membrane localization.

In various embodiments of the method, the assay may take the form of a yeast growth assay, an Aequorin assay, a Luciferase assay, a mitogenesis assay, a MAP Kinase activity assay, as well as other binding or function-based assays of kinase activity that are generally known in the art. In several of these embodiments, the invention includes any of the receptor and non-receptor protein tyrosine kinases, receptor and non-receptor protein phosphatases, polypeptides containing *SRC* homology 2 and 3 domains, phosphotyrosine binding proteins (*SRC* homology 2 (SH2) and phosphotyrosine binding (PTB and PH) domain containing proteins), proline-rich binding proteins (SH3 domain containing proteins), GTPases, phosphodiesterases, phospholipases, prolyl isomerases, proteases, Ca2+ binding proteins, cAMP binding proteins, guanyl cyclases, adenylyl cyclases, NO generating proteins, nucleotide exchange factors, and transcription factors. Biological activities of kinases according to the invention include, but are not limited to, the binding of a natural or a synthetic

-45-

ligand, as well as any one of the functional activities of kinases known in the art.
Non-limiting examples of kinase activities include transmembrane signaling of
various forms, which may involve kinase binding interactions and/or the exertion of
an influence over signal transduction.

The modulators of the invention exhibit a variety of chemical structures, which can be
generally grouped into mimetics of natural kinase ligands, and peptide and non-
peptide allosteric effectors of kinases. The invention does not restrict the sources for
suitable modulators, which may be obtained from natural sources such as plant,
animal or mineral extracts, or non-natural sources such as small molecule libraries,
including the products of combinatorial chemical approaches to library construction,
and peptide libraries.

The use of cDNAs encoding kinases in drug discovery programs is well-known;
assays capable of testing thousands of unknown compounds per day in high-
throughput screens (HTSs) are thoroughly documented. The literature is replete with
examples of the use of radiolabelled ligands in HTS binding assays for drug discovery
(see Williams, *Medicinal Research Reviews*, 1991, *11*, 147-184.; Sweetnam, *et al., J.
Natural Products*, 1993, *56*, 441-455 for review). Recombinant proteins are preferred
for binding assay HTS because they allow for better specificity (higher relative
purity), provide the ability to generate large amounts of material, and can be used in a
broad variety of formats (see Hodgson, *Bio/Technology*, 1992, *10*, 973-980; each of
which is incorporated herein by reference in its entirety).

A variety of heterologous systems is available for functional expression of
recombinant proteins that are well known to those skilled in the art. Such systems
include bacteria (Strosberg, *et al., Trends in Pharmacological Sciences*, 1992, *13*, 95-
98), yeast (Pausch, *Trends in Biotechnology*, 1997, *15*, 487-494), several kinds of
insect cells (Vanden Broeck, *Int. Rev. Cytology*, 1996, *164*, 189-268), amphibian cells
(Jayawickreme *et al., Current Opinion in Biotechnology*, 1997, *8*, 629-634) and
several mammalian cell lines (CHO, HEK293, COS, etc.; see Gerhardt, *et al., Eur. J.
Pharmacology*, 1997, *334*, 1-23). These examples do not preclude the use of other

-46-

possible cell expression systems, including cell lines obtained from nematodes (PCT application WO 98/37177).

An expressed kinase can be used for HTS binding assays in conjunction with its defined ligand, in this case the corresponding peptide that activates it. The identified peptide is labeled with a suitable radioisotope, including, but not limited to, ^{125}I, ^{3}H, ^{35}S or ^{32}P, by methods that are well known to those skilled in the art. Alternatively, the peptides may be labeled by well-known methods with a suitable fluorescent derivative (Baindur, *et al.*, *Drug Dev. Res.*, 1994, 33, 373-398; Rogers, *Drug Discovery Today*, 1997, 2, 156-160). Radioactive ligand specifically bound to the receptor in membrane preparations made from the cell line expressing the recombinant protein can be detected in HTS assays in one of several standard ways, including filtration of the receptor-ligand complex to separate bound ligand from unbound ligand (Williams, *Med. Res. Rev.*, 1991, 11, 147-184.; Sweetnam, *et al.*, *J. Natural Products*, 1993, 56, 441-455). Alternative methods include a scintillation proximity assay (SPA) or a FlashPlate format in which such separation is unnecessary (Nakayama, *Cur. Opinion Drug Disc. Dev.*, 1998, 1, 85-91 Bossé, *et al.*, *J. Biomolecular Screening*, 1998, 3, 285-292.). Binding of fluorescent ligands can be detected in various ways, including fluorescence energy transfer (FRET), direct spectrophotofluorometric analysis of bound ligand, or fluorescence polarization (Rogers, *Drug Discovery Today*, 1997, 2, 156-160; Hill, *Cur. Opinion Drug Disc. Dev.*, 1998, 1, 92-97).

The kinases and natural binding partners required for functional expression of heterologous kinase polypeptides can be native constituents of the host cell or can be introduced through well-known recombinant technology. The kinase polypeptides can be intact or chimeric. The kinase activation results in the stimulation or inhibition of other native proteins, events that can be linked to a measurable response.

Examples of such biological responses include, but are not limited to, the following: the ability to survive in the absence of a limiting nutrient in specifically engineered yeast cells (Pausch, *Trends in Biotechnology*, 1997, 15, 487-494); changes in intracellular Ca^{2+} concentration as measured by fluorescent dyes (Murphy, *et al.*, *Cur.*

-47-

Opinion Drug Disc. Dev., 1998, *1*, 192-199), cell cycle, apoptosis, and growth.
Fluorescence changes can also be used to monitor ligand-induced changes in
membrane potential or intracellular pH; an automated system suitable for HTS has
been described for these purposes (Schroeder, *et al., J. Biomolecular Screening*, 1996,
1, 75-80).

The invention contemplates a multitude of assays to screen and identify inhibitors of
ligand binding to kinase polypeptides. In one example, the kinase polypeptide is
immobilized and interaction with a binding partner is assessed in the presence and
absence of a candidate modulator such as an inhibitor compound. In another
example, interaction between the kinase polypeptide and its binding partner is
assessed in a solution assay, both in the presence and absence of a candidate inhibitor
compound. In either assay, an inhibitor is identified as a compound that decreases
binding between the kinase polypeptide and its natural binding partner. Another
contemplated assay involves a variation of the di-hybrid assay wherein an inhibitor of
protein/protein interactions is identified by detection of a positive signal in a
transformed or transfected host cell, as described in PCT publication number WO
95/20652, published August 3, 1995 and is included by reference herein including any
figures, tables, or drawings.

Candidate modulators contemplated by the invention include compounds selected
from libraries of either potential activators or potential inhibitors. There are a number
of different libraries used for the identification of small molecule modulators,
including: (1) chemical libraries, (2) natural product libraries, and (3) combinatorial
libraries comprised of random peptides, oligonucleotides or organic molecules.
Chemical libraries consist of random chemical structures, some of which are analogs
of known compounds or analogs of compounds that have been identified as "hits" or
"leads" in other drug discovery screens, while others are derived from natural
products, and still others arise from non-directed synthetic organic chemistry. Natural
product libraries are collections of microorganisms, animals, plants, or marine
organisms which are used to create mixtures for screening by: (1) fermentation and
extraction of broths from soil, plant or marine microorganisms or (2) extraction of

-48-

plants or marine organisms. Natural product libraries include polyketides, non-ribosomal peptides, and variants (non-naturally occurring) thereof. For a review, see *Science* 282: 63-68 (1998). Combinatorial libraries are composed of large numbers of peptides, oligonucleotides, or organic compounds as a mixture. These libraries are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning, or proprietary synthetic methods. Of particular interest are non-peptide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8: 701-707 (1997). Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to modulate activity.

Still other candidate inhibitors contemplated by the invention can be designed and include soluble forms of binding partners, as well as such binding partners as chimeric, or fusion, proteins. A "binding partner" as used herein broadly encompasses both natural binding partners as described above as well as chimeric polypeptides, peptide modulators other than natural ligands, antibodies, antibody fragments, and modified compounds comprising antibody domains that are immunospecific for the expression product of the identified kinase gene.

Other assays may be used to identify specific peptide ligands of a kinase polypeptide, including assays that identify ligands of the target protein through measuring direct binding of test ligands to the target protein, as well as assays that identify ligands of target proteins through affinity ultrafiltration with ion spray mass spectroscopy/HPLC methods or other physical and analytical methods. Alternatively, such binding interactions are evaluated indirectly using the yeast two-hybrid system described in Fields *et al., Nature,* 340: 245-246 (1989), and Fields *et al., Trends in Genetics,* 10: 286-292 (1994), both of which are incorporated herein by reference. The two-hybrid system is a genetic assay for detecting interactions between two proteins or polypeptides. It can be used to identify proteins that bind to a known protein of interest, or to delineate domains or residues critical for an interaction. Variations on

-49-

this methodology have been developed to clone genes that encode DNA binding proteins, to identify peptides that bind to a protein, and to screen for drugs. The two-hybrid system exploits the ability of a pair of interacting proteins to bring a transcription activation domain into close proximity with a DNA binding domain that binds to an upstream activation sequence (UAS) of a reporter gene, and is generally performed in yeast. The assay requires the construction of two hybrid genes encoding (1) a DNA-binding domain that is fused to a first protein and (2) an activation domain fused to a second protein. The DNA-binding domain targets the first hybrid protein to the UAS of the reporter gene; however, because most proteins lack an activation domain, this DNA-binding hybrid protein does not activate transcription of the reporter gene. The second hybrid protein, which contains the activation domain, cannot by itself activate expression of the reporter gene because it does not bind the UAS. However, when both hybrid proteins are present, the noncovalent interaction of the first and second proteins tethers the activation domain to the UAS, activating transcription of the reporter gene. For example, when the first protein is a kinase gene product, or fragment thereof, that is known to interact with another protein or nucleic acid, this assay can be used to detect agents that interfere with the binding interaction. Expression of the reporter gene is monitored as different test agents are added to the system. The presence of an inhibitory agent results in lack of a reporter signal.

When the function of the kinase polypeptide gene product is unknown and no ligands are known to bind the gene product, the yeast two-hybrid assay can also be used to identify proteins that bind to the gene product. In an assay to identify proteins that bind to a kinase polypeptide, or fragment thereof, a fusion polynucleotide encoding both a kinase polypeptide (or fragment) and a UAS binding domain (i.e., a first protein) may be used. In addition, a large number of hybrid genes each encoding a different second protein fused to an activation domain are produced and screened in the assay. Typically, the second protein is encoded by one or more members of a total cDNA or genomic DNA fusion library, with each second protein coding region being fused to the activation domain. This system is applicable to a wide variety of proteins, and it is not even necessary to know the identity or function of the second binding protein. The system is highly sensitive and can detect interactions not

revealed by other methods; even transient interactions may trigger transcription to produce a stable mRNA that can be repeatedly translated to yield the reporter protein.

Other assays may be used to search for agents that bind to the target protein. One such screening method to identify direct binding of test ligands to a target protein is described in U.S. Patent No. 5,585,277, incorporated herein by reference. This method relies on the principle that proteins generally exist as a mixture of folded and unfolded states, and continually alternate between the two states. When a test ligand binds to the folded form of a target protein (*i.e.*, when the test ligand is a ligand of the target protein), the target protein molecule bound by the ligand remains in its folded state. Thus, the folded target protein is present to a greater extent in the presence of a test ligand which binds the target protein, than in the absence of a ligand. Binding of the ligand to the target protein can be determined by any method which distinguishes between the folded and unfolded states of the target protein. The function of the target protein need not be known in order for this assay to be performed. Virtually any agent can be assessed by this method as a test ligand, including, but not limited to, metals, polypeptides, proteins, lipids, polysaccharides, polynucleotides and small organic molecules.

Another method for identifying ligands of a target protein is described in Wieboldt *et al.*, *Anal. Chem.*, 69: 1683-1691 (1997), incorporated herein by reference. This technique screens combinatorial libraries of 20-30 agents at a time in solution phase for binding to the target protein. Agents that bind to the target protein are separated from other library components by simple membrane washing. The specifically selected molecules that are retained on the filter are subsequently liberated from the target protein and analyzed by HPLC and pneumatically assisted electrospray (ion spray) ionization mass spectroscopy. This procedure selects library components with the greatest affinity for the target protein, and is particularly useful for small molecule libraries.

In preferred embodiments of the invention, methods of screening for compounds which modulate kinase activity comprise contacting test compounds with kinase polypeptides and assaying for the presence of a complex between the compound and

-51-

the kinase polypeptide. In such assays, the ligand is typically labelled. After suitable incubation, free ligand is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability of the particular compound to bind to the kinase polypeptide.

In another embodiment of the invention, high throughput screening for compounds having suitable binding affinity to kinase polypeptides is employed. Briefly, large numbers of different small peptide test compounds are synthesised on a solid substrate. The peptide test compounds are contacted with the kinase polypeptide and washed. Bound kinase polypeptide is then detected by methods well known in the art. Purified polypeptides of the invention can also be coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies can be used to capture the protein and immobilize it on the solid support.

Other embodiments of the invention comprise using competitive screening assays in which neutralizing antibodies capable of binding a polypeptide of the invention specifically compete with a test compound for binding to the polypeptide. In this manner, the antibodies can be used to detect the presence of any peptide that shares one or more antigenic determinants with a kinase polypeptide. Radiolabeled competitive binding studies are described in A.H. Lin *et al. Antimicrobial Agents and Chemotherapy*, 1997, vol. 41, no. 10. pp. 2127-2131, the disclosure of which is incorporated herein by reference in its entirety.

In another aspect, the invention provides methods for treating a disease by administering to a patient in need of such treatment a substance that modulates the activity of a kinase polypeptide selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 , as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein. Preferably the disease is selected from the group consisting of cancers, immune-elated diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, skin or kidney; central or peripheral

-52-

nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, hypertension, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, bone disorders, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

In preferred embodiments, the invention provides methods for treating or preventing a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 , as well as the full-length polypeptide thereof, or a portion of any of these sequences that retains functional activity, as described herein. Preferably, the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's,

Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial-organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

Substances useful for treatment of kinase-related disorders or diseases preferably show positive results in one or more *in vitro* assays for an activity corresponding to treatment of the disease or disorder in question (Examples of such assays are provided in the references in section VI, below; and in Example 7, herein). Examples of substances that can be screened for favorable activity are provided and referenced in section VI, below. The substances that modulate the activity of the kinases preferably include, but are not limited to, antisense oligonucleotides and inhibitors of protein kinases, as determined by methods and screens referenced in section VI and Example 7, below.

The term "preventing" refers to decreasing the probability that an organism contracts or develops an abnormal condition.

The term "treating" refers to having a therapeutic effect and at least partially alleviating or abrogating an abnormal condition in the organism.

The term "therapeutic effect" refers to the inhibition or activation factors causing or contributing to the abnormal condition. A therapeutic effect relieves to some extent one or more of the symptoms of the abnormal condition. In reference to the treatment of abnormal conditions, a therapeutic effect can refer to one or more of the following: (a) an decrease in the proliferation, growth, and/or differentiation of cells; (b) inhibition (*i.e.*, slowing or stopping) of cell death; (c) inhibition of degeneration; (d)

-54-

relieving to some extent one or more of the symptoms associated with the abnormal condition; and (e) enhancing the function of the affected population of cells. Compounds demonstrating efficacy against abnormal conditions can be identified as described herein.

The term "abnormal condition" refers to a function in the cells or tissues of an organism that deviates from their normal functions in that organism. An abnormal condition can relate to cell proliferation, cell differentiation, or cell survival.

Abnormal cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

Abnormal differentiation conditions include, but are not limited to neurodegenerative disorders, slow wound healing rates, and slow tissue grafting healing rates.

Abnormal cell survival conditions relate to conditions in which programmed cell death (apoptosis) pathways are activated or abrogated. A number of protein kinases are associated with the apoptosis pathways. Aberrations in the function of any one of the protein kinases could lead to cell immortality or premature cell death.

The term "aberration," in conjunction with the function of a kinase in a signal transduction process, refers to a kinase that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type protein kinase activity, mutated such that it can no longer interact with a natural binding partner, is no longer modified by another protein kinase or protein phosphatase, or no longer interacts with a natural binding partner.

The term "administering" relates to a method of incorporating a compound into cells or tissues of an organism. The abnormal condition can be prevented or treated when the cells or tissues of the organism exist within the organism or outside of the organism. Cells existing outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, dermal,

injection, and aerosol applications. For cells outside of the organism, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques, and carrier techniques.

The abnormal condition can also be prevented or treated by administering a compound to a group of cells having an aberration in a signal transduction pathway to an organism. The effect of administering a compound on organism function can then be monitored. The organism is preferably a mammal. The organism also is preferably a mouse, rat, rabbit, guinea pig, dog, cat, horse, pig, sheep, or goat, more preferably a monkey or ape, and most preferably a human.

In another aspect, the invention features methods for detection of a kinase polypeptide in a sample as a diagnostic tool for diseases or disorders, wherein the method comprises the steps of: (a) contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a kinase polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 , said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and (b) detecting the presence or amount of the probe: target region hybrid as an indication of the disease.

In preferred embodiments of the invention, the disease or disorder is selected from the group consisting of Preferably the disease is selected from the group consisting of cancers, immune-elated diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, skin or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis,

-56-

and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1,
HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic
disorders including Diabetes and obesity and their related syndromes, among others;
cardiovascular disorders including reperfusion restenosis, hypertension, coronary
thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis;
ocular disease including glaucoma, retinopathy, and macular degeneration;
inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel
disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma,
osteoarthritis, bone disorders, psoriasis, atherosclerosis, rhinitis, autoimmunity, and
organ transplant rejection.

The kinase "target region" is the nucleotide base sequence selected from the group
consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66, or the
corresponding full-length sequences, a functional derivative thereof, or a fragment
thereof, to which the nucleic acid probe will specifically hybridize. Specific
hybridization indicates that in the presence of other nucleic acids the probe only
hybridizes detectably with the kinase of the invention's target region. Putative target
regions can be identified by methods well known in the art consisting of alignment
and comparison of the most closely related sequences in the database.

In preferred embodiments the nucleic acid probe hybridizes to a kinase target region
encoding at least 6, 12, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino
acids of a sequence selected from the group consisting of those set forth in SEQ ID
NO: 67 through 132 , or the corresponding full-length amino acid sequence, a portion
of any of these sequences that retains functional activity, as described herein, or a
functional derivative thereof. Hybridization conditions should be such that
hybridization occurs only with the kinase genes in the presence of other nucleic acid
molecules. Under stringent hybridization conditions only highly complementary
nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization
of nucleic acids having more than 1 or 2 mismatches out of 20 contiguous
nucleotides. Such conditions are defined *supra*.

The diseases for which detection of kinase genes in a sample could be diagnostic include diseases in which kinase nucleic acid (DNA and/or RNA) is amplified in comparison to normal cells. By "amplification" is meant increased numbers of kinase DNA or RNA in a cell compared with normal cells. In normal cells, kinases are typically found as single copy genes. In selected diseases, the chromosomal location of the kinase genes may be amplified, resulting in multiple copies of the gene, or amplification. Gene amplification can lead to amplification of kinase RNA, or kinase RNA can be amplified in the absence of kinase DNA amplification.

"Amplification" as it refers to RNA can be the detectable presence of kinase RNA in cells, since in some normal cells there is no basal expression of kinase RNA. In other normal cells, a basal level of expression of kinase exists, therefore in these cases amplification is the detection of at least 1-2-fold, and preferably more, kinase RNA, compared to the basal level.

The diseases that could be diagnosed by detection of kinase nucleic acid in a sample preferably include cancers or other diseases described herein. The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample that is compatible with the method utilized.

The invention also features a method for detection of a kinase polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein the method comprises: (a) comparing a nucleic acid target region encoding the kinase polypeptide in a sample, where the kinase polypeptide has an amino acid sequence selected from the group consisting those set forth in SEQ ID NO: 67 through SEQ ID NO: 132, or one or more fragments thereof, with a control nucleic acid target region encoding the kinase polypeptide, or one or more fragments thereof; and (b) detecting differences in sequence or amount between the target region and the control target region, as an indication of the disease or disorder. Preferably the disease is selected from the group

-58-

consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

The term "comparing" as used herein refers to identifying discrepancies between the nucleic acid target region isolated from a sample, and the control nucleic acid target region. The discrepancies can be in the nucleotide sequences, e.g. insertions, deletions, or point mutations, or in the amount of a given nucleotide sequence. Methods to determine these discrepancies in sequences are well-known to one of ordinary skill in the art. The "control" nucleic acid target region refers to the sequence or amount of the sequence found in normal cells, e.g. cells that are not diseased as discussed previously.

The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

-59-

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the nucleotide sequences for human protein kinases oriented in a 5' to 3' direction (SEQ ID NO: 1-66).

Figure 2 shows the amino acid sequences for the human protein kinases encoded by SEQ ID No. 1 and 2 in the direction of translation (SEQ ID NO: 67 through 132). If a predicted stop codons is within the coding region, it is indicated by an 'x.'

DETAILED DESCRIPTION OF THE INVENTION

The invention provides, *inter alia,* protein kinase and kinase-like genes, as well as fragments thereof, which have been identified in genomic databases. In part, the invention provides nucleic acid molecules that are capable of encoding polypeptides having a kinase or kinase-like activity. By reference to Tables 1 though 6, below, genes of the invention can be better understood. The invention additionally provides a number of different embodiments, such as those described below.

Nucleic Acids

Associations of chromosomal localizations for mapped genes with amplicons implicated in cancer are based on literature searches (PubMed http: //www.ncbi.nlm.nih.gov/entrez/query.fcgi), OMIM searches (Online Mendelian Inheritance in Man, http: //www.ncbi.nlm.nih.gov/Omim/searchomim.html) and the comprehensive database of cancer amplicons maintained by Knuutila, et al. (Knuutila, et al., DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123, 1998. http: //www.helsinki.fi/~lgl www/CMG.html).

For single nucleotide polymorphisms, an accession number is given if the SNP is documented in dbSNP (the database of single nucleotide polymorphisms) maintained at NCBI (http: //www.ncbi.nlm.nih.gov/SNP/index.html). The accession number for SNP can be used to retrieve the full SNP-containing sequence from this site.

All of the sequences are derived from human DNA, with the exception of Pak4, which is from Mus musculus.

NUCLEIC ACID PROBES, METHODS, AND KITS FOR DETECTION OF KINASES

The invention additionally provides nucleic acid probes and uses therefor. A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (cf. "Molecular Cloning: A Laboratory Manual," second edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications," Academic Press, Michael, et al., eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art ("Molecular Cloning: A Laboratory Manual," 1989, supra). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates

-61-

such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said probe bound to said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or steptavidin). Preferably, the kit further comprises instructions for use.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are

-62-

not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

CATEGORIZATION OF THE POLYPEPTIDES ACCORDING TO THE INVENTION

For a number of protein kinases of the invention, there is provided a classification of the protein class and family to which it belongs, a summary of non-catalytic protein motifs, as well as a chromosomal location, which provides information on function, regulation and/or therapeutic utility for each of the proteins. Amplification of chromosomal region can be associated with various cancers. For amplicons discussed in this application, the source of information was Knuutila, et al (Knuutila S, Björkqvist A-M, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H, El-Rifai W, Hemmer S, Wasenius V-M, Vidgren V & Zhu Y: DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123, 1998. http: //www.helsinki.fi/~lgl_www/CMG.html).

The kinase classification and protein domains often reflect pathways, cellular roles, or mechanisms of up- or down-stream regulation. Also disease-relevant genes often occur in families of related genes. For example, if one member of a kinase family functions as an oncogene, a tumor suppressor, or has been found to be disrupted in an immune, neurologic, cardiovascular, or metabolic disorder, frequently other family members may play a similar role.

Chromosomal location can identify candidate targets for a tumor amplicon or a tumor-suppressor locus. Summaries of prevalent tumor amplicons are available in the

-63-

literature, and can identify tumor types to experimentally be confirmed to contain amplified copies of a kinase gene which localizes to an adjacent region.

As described herein, the polypeptides of the present invention can be classified. The salient features related to the biological and clinical implications of these different groups are described hereafter in more general terms.

A more specific characterization of the polypeptides of the invention, including potential biological and clinical implications, is provided, e.g., in EXAMPLES 2a and 2b.

CLASSIFICATION OF POLYPEPTIDES EXHIBITING KINASE ACTIVITY

The classification of the polypeptides described in this application is found in Tables 1 and 2. The present application describes members of the following superfamilies: protein kinase, lipid kinase, atypical protein kinase. The present application also describes members of the following groups: CAMK Group, CKI (or CK1) Group, CMGC Group, STE Group, TK Group, DAG (diacylglycerol) Group, BRD Group. ***

Potential biological and clinical implications of these novel kinases are described below.

THERAPEUTIC METHODS ACCORDING TO THE INVENTION:

Diagnostics:

The invention provides methods for detecting a polypeptide in a sample as a diagnostic tool for diseases or disorders, wherein the method comprises the steps of: (a) contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a polypeptide selected from the group consisting of SEQ ID NO: 67 through 132, said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and (b) detecting the presence or amount of the probe: target region hybrid as an indication of the disease.

-64-

In preferred embodiments of the invention, the disease or disorder is selected from the group consisting of rheumatoid arthritis, atherosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, metabolic disorder including diabetes, reproductive disorders including infertility, and cancer.

Hybridization conditions should be such that hybridization occurs only with the genes in the presence of other nucleic acid molecules. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides. Such conditions are defined supra.

The diseases for which detection of genes in a sample could be diagnostic include diseases in which nucleic acid (DNA and/or RNA) is amplified in comparison to normal cells. By "amplification" is meant increased numbers of DNA or RNA in a cell compared with normal cells.

"Amplification" as it refers to RNA can be the detectable presence of RNA in cells, since in some normal cells there is no basal expression of RNA. In other normal cells, a basal level of expression exists, therefore in these cases amplification is the detection of at least 1-2-fold, and preferably more, compared to the basal level.

The diseases that could be diagnosed by detection of nucleic acid in a sample preferably include cancers. The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample that is compatible with the method utilized.

Antibodies, Hybridomas, Methods of Use and Kits for Detection of Kinases

The present invention relates to an antibody having binding affinity to a kinase of the invention. The polypeptide may have the amino acid sequence selected from the

-65-

group consisting of those set forth in SEQ ID NO: 67 through 132 , or a functional derivative thereof, or at least 9 contiguous amino acids thereof (preferably, at least 20, 30, 35, or 40 contiguous amino acids thereof).

The present invention also relates to an antibody having specific binding affinity to a kinase of the invention. Such an antibody may be isolated by comparing its binding affinity to a kinase of the invention with its binding affinity to other polypeptides. Those which bind selectively to a kinase of the invention would be chosen for use in methods requiring a distinction between a kinase of the invention and other polypeptides. Such methods could include, but should not be limited to, the analysis of altered kinase expression in tissue containing other polypeptides.

The kinases of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for studying DNA/protein interaction.

The kinases of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide could be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the present invention may be generated using one of the procedures known in the art such as chimerization or CDR grafting.

The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth et al., J. Immunol. Methods 35: 1-21, 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for

-66-

immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or β-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Agl4 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al., Exp. Cell Res.* 175: 109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass are determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," *supra*, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, *see* Stemberger *et al., J. Histochem. Cytochem.* 18: 315, 1970; Bayer *et al., Meth. Enzym.* 62: 308, 1979; Engval *et al., Immunol.* 109: 129, 1972; Goding, *J. Immunol. Meth.* 13: 215, 1976. The labeled antibodies of the present invention can

-67-

be used for *in vitro, in vivo,* and *in situ* assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.,* "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby *et al., Meth. Enzym.* 34, Academic Press, N.Y., 1974). The immobilized antibodies of the present invention can be used for *in vitro, in vivo,* and *in situ* assays as well as in immunochromotography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed herein with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides (Hurby *et al.,* "Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307, 1992; Kaspczak *et al., Biochemistry* 28: 9230-9238, 1989).

Anti-peptide peptides can be generated by replacing the basic amino acid residues found in the peptide sequences of the kinases of the invention with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

The present invention also encompasses a method of detecting a kinase polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying

-68-

whether the antibody binds to the test sample. Altered levels of a kinase of the invention in a sample as compared to normal levels may indicate disease.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion-based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard ("An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands, 1986), Bullock *et al.* ("Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985), Tijssen ("Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test samples used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can readily be adapted in order to obtain a sample which is testable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may comprise: (i) a first container means containing an above-described antibody, and (ii) second container means containing a conjugate comprising a binding partner of the antibody and a label. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

Isolation of Compounds Capable of Interacting with Kinases

The present invention also relates to a method of detecting a compound capable of binding to a kinase of the invention comprising incubating the compound with a kinase of the invention and detecting the presence of the compound bound to the kinase. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts.

The present invention also relates to a method of detecting an agonist or antagonist of kinase activity or kinase binding partner activity comprising incubating cells that produce a kinase of the invention in the presence of a compound and detecting changes in the level of kinase activity or kinase binding partner activity. The compounds thus identified would produce a change in activity indicative of the presence of the compound. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts. Once the compound is identified it can be isolated using techniques well known in the art.

Modulating polypeptide activity:

The invention additionally provides methods for treating a disease or abnormal condition by administering to a patient in need of such treatment a substance that modulates the activity of a polypeptide selected from the group consisting of SEQ ID NO: 67 through 132. Preferably, the disease is selected from the group consisting of rheumatoid arthritis, atherosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, metabolic and reproductive disorders, and cancer.

-70-

Substances useful for treatment of disorders or diseases preferably show positive results in one or more assays for an activity corresponding to treatment of the disease or disorder in question Substances that modulate the activity of the polypeptides preferably include, but are not limited to, antisense oligonucleotides and inhibitors of protein kinases.

The term "preventing" refers to decreasing the probability that an organism contracts or develops an abnormal condition.

The term "treating" refers to having a therapeutic effect and at least partially alleviating or abrogating an abnormal condition in the organism.

The term "therapeutic effect" refers to the inhibition or activation factors causing or contributing to the abnormal condition. A therapeutic effect relieves to some extent one or more of the symptoms of the abnormal condition. In reference to the treatment of abnormal conditions, a therapeutic effect can refer to one or more of the following: (a) a decrease in the proliferation, growth, and/or differentiation of cells; (b) inhibition (, slowing or stopping) of cell death; (c) inhibition of degeneration; (d) relieving to some extent one or more of the symptoms associated with the abnormal condition; and (e) enhancing the function of the affected population of cells. Compounds demonstrating efficacy against abnormal conditions can be identified as described herein.

The term "abnormal condition" refers to a function in the cells or tissues of an organism that deviates from their normal functions in that organism. An abnormal condition can relate to cell proliferation, cell differentiation or cell survival. An abnormal condition may also include irregularities in cell cycle progression, i.e., irregularities in normal cell cycle progression through mitosis and meiosis.

Abnormal cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

Abnormal differentiation conditions include, but are not limited to, neurodegenerative disorders, slow wound healing rates, and slow tissue grafting healing rates.

Abnormal cell survival conditions may also relate to conditions in which programmed cell death (apoptosis) pathways are activated or abrogated. A number of protein kinases are associated with the apoptosis pathways. Aberrations in the function of any one of the protein kinases could lead to cell immortality or premature cell death.

The term "aberration," in conjunction with the function of a kinase in a signal transduction process, refers to a kinase that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type protein kinase activity, mutated such that it can no longer interact with a natural binding partner, is no longer modified by another protein kinase or protein phosphatase, or no longer interacts with a natural binding partner.

The term "administering" relates to a method of incorporating a compound into cells or tissues of an organism. The abnormal condition can be prevented or treated when the cells or tissues of the organism exist within the organism or outside of the organism. Cells existing outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, dermal, injection, and aerosol applications. For cells outside of the organism, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques and carrier techniques.

The abnormal condition can also be prevented or treated by administering a compound to a group of cells having an aberration in a signal transduction pathway to an organism. The effect of administering a compound on organism function can then be monitored. The organism is preferably a mouse, rat, rabbit, guinea pig or goat, more preferably a monkey or ape, and most preferably a human.

The present invention also encompasses a method of agonizing (stimulating) or antagonizing kinase associated activity in a mammal comprising administering to said mammal an agonist or antagonist to a kinase of the invention in an amount sufficient

to effect said agonism or antagonism. A method of treating diseases in a mammal with an agonist or antagonist of the activity of one of the kinases of the invention comprising administering the agonist or antagonist to a mammal in an amount sufficient to agonize or antagonize kinase-associated functions is also encompassed in the present application.

In an effort to discover novel treatments for diseases, biomedical researchers and chemists have designed, synthesized, and tested molecules that inhibit the function of protein kinases. Some small organic molecules form a class of compounds that modulate the function of protein kinases. Examples of molecules that have been reported to inhibit the function of some protein kinases include, but are not limited to, bis monocyclic, bicyclic or heterocyclic aryl compounds (PCT WO 92/20642, published November 26, 1992 by Maguire *et al.*), vinylene-azaindole derivatives (PCT WO 94/14808, published July 7, 1994 by Ballinari *et al.*), 1-cyclopropyl-4-pyridyl-quinolones (U.S. Patent No. 5,330,992), styryl compounds (U.S. Patent No. 5,217,999), styryl-substituted pyridyl compounds (U.S. Patent No. 5,302,606), certain quinazoline derivatives (EP Application No. 0 566 266 A1), seleoindoles and selenides (PCT WO 94/03427, published February 17, 1994 by Denny *et al.*), tricyclic polyhydroxylic compounds (PCT WO 92/21660, published December 10, 1992 by Dow), and benzylphosphonic acid compounds (PCT WO 91/15495, published October 17, 1991 by Dow *et al*).

Compounds that can traverse cell membranes and are resistant to acid hydrolysis are potentially advantageous as therapeutics as they can become highly bioavailable after being administered orally to patients. However, many of these protein kinase inhibitors only weakly inhibit the function of protein kinases. In addition, many inhibit a variety of protein kinases and will therefore cause multiple side-effects as therapeutics for diseases.

Some indolinone compounds, however, form classes of acid resistant and membrane permeable organic molecules. WO 96/22976 (published August 1, 1996 by Ballinari *et al.*) describes hydrosoluble indolinone compounds that harbor tetralin, naphthalene, quinoline, and indole substituents fused to the oxindole ring. These bicyclic

-73-

substituents are in turn substituted with polar moieties including hydroxylated alkyl, phosphate, and ether moieties. U.S. Patent Application Serial Nos. 08/702,232, filed August 23, 1996, entitled "Indolinone Combinatorial Libraries and Related Products and Methods for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 221/187) and 08/485,323, filed June 7, 1995, entitled "Benzylidene-Z-Indoline Compounds for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 223/298) and International Patent Publications WO 96/40116, published December 19, 1996 by Tang, *et al.*, and WO 96/22976, published August 1, 1996 by Ballinari *et al.*, all of which are incorporated herein by reference in their entirety, including any drawings, figures, or tables, describe indolinone chemical libraries of indolinone compounds harboring other bicyclic moieties as well as monocyclic moieties fused to the oxindole ring. Applications 08/702,232, filed August 23, 1996, entitled "Indolinone Combinatorial Libraries and Related Products and Methods for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 221/187), 08/485,323, filed June 7, 1995, entitled "Benzylidene-Z-Indoline Compounds for the Treatment of Disease" by Tang *et al.* (Lyon & Lyon Docket No. 223/298), and WO 96/22976, published August 1, 1996 by Ballinari *et al.* teach methods of indolinone synthesis, methods of testing the biological activity of indolinone compounds in cells, and inhibition patterns of indolinone derivatives.

Other examples of substances capable of modulating kinase activity include, but are not limited to, tyrphostins, quinazolines, quinoxolines, and quinolines. The quinazolines, tyrphostins, quinolines, and quinoxolines referred to above include well known compounds such as those described in the literature. For example, representative publications describing quinazolines include Barker *et al.*, EPO Publication No. 0 520 722 A1; Jones *et al.*, U.S. Patent No. 4,447,608; Kabbe *et al.*, U.S. Patent No. 4,757,072; Kaul and Vougioukas, U.S. Patent No. 5,316,553; Kreighbaum and Comer, U.S. Patent No. 4,343,940; Pegg and Wardleworth, EPO Publication No. 0 562 734 A1; Barker *et al.*, (1991) *Proc. of Am. Assoc. for Cancer Research* 32: 327; Bertino, J.R., (1979) *Cancer Research* 3: 293-304; Bertino, J.R., (1979) *Cancer Research* 9(2 part 1): 293-304; Curtin *et al.*, (1986) *Br. J. Cancer* 53: 361-368; Fernandes *et al.*, (1983) *Cancer Research* 43: 1117-1123 ; Ferris *et al. J.*

Org. Chem. 44(2): 173-178; Fry *et al.*, (1994) *Science* 265: 1093-1095; Jackman *et al.*, (1981) *Cancer Research* 51: 5579-5586; Jones *et al. J. Med. Chem.* 29(6): 1114-1118; Lee and Skibo, (1987) *Biochemistry* 26(23): 7355-7362; Lemus *et al.*, (1989) *J. Org. Chem.* 54: 3511-3518; Ley and Seng, (1975) *Synthesis* 1975: 415-522; Maxwell *et al.*, (1991) *Magnetic Resonance in Medicine* 17: 189-196 ; Mini *et al.*, (1985) *Cancer Research* 45: 325-330; Phillips and Castle, J. (1980) *Heterocyclic Chem.* 17(19): 1489-1596; Reece *et al.*, (1977) *Cancer Research* 47(11): 2996-2999; Sculier *et al.*, (1986) *Cancer Immunol. and Immunother.* 23, A65; Sikora *et al.*, (1984) *Cancer Letters* 23: 289-295; Sikora *et al.*, (1988) *Analytical Biochem.* 172: 344-355; all of which are incorporated herein by reference in their entirety, including any drawings.

Quinoxaline is described in Kaul and Vougioukas, U.S. Patent No. 5,316,553, incorporated herein by reference in its entirety, including any drawings.

Quinolines are described in Dolle *et al.*, (1994) *J. Med. Chem.* 37: 2627-2629; MaGuire, J. (1994) *Med. Chem.* 37: 2129-2131; Burke *et al.*, (1993) *J. Med. Chem.* 36: 425-432 ; and Burke *et al.* (1992) *BioOrganic Med. Chem. Letters* 2: 1771-1774, all of which are incorporated by reference in their entirety, including any drawings.

Tyrphostins are described in Allen *et al.*, (1993) *Clin. Exp. Immunol.* 91: 141-156; Anafi *et al.*, (1993) *Blood* 82: 12, 3524-3529; Baker *et al.*, (1992) *J. Cell Sci.* 102: 543-555; Bilder *et al.*, (1991) *Amer. Physiol. Soc.* pp. 6363-6143: C721-C730; Brunton *et al.*, (1992) *Proceedings of Amer. Assoc. Cancer Rsch.* 33: 558; Bryckaert *et al.*, (1992) *Exp. Cell Research* 199: 255-261; Dong *et al.*, (1993) *J. Leukocyte Biology* 53: 53-60; Dong *et al.*, (1993) *J. Immunol.* 151(5): 2717-2724; Gazit *et al.*, (1989) *J. Med. Chem.* 32, 2344-2352; Gazit *et al.*, (1993) *J. Med. Chem.* 36: 3556-3564; Kaur *et al.*, (1994) *Anti-Cancer Drugs* 5: 213-222; King *et al.*, (1991) *Biochem. J.* 275: 413-418; Kuo *et al.*, (1993) *Cancer Letters* 74: 197-202; Levitzki, A., (1992) *The FASEB J.* 6: 3275-3282; Lyall *et al.*, (1989) *J. Biol. Chem.* 264: 14503-14509; Peterson *et al.*, (1993) *The Prostate* 22: 335-345; Pillemer *et al.*, (1992) *Int. J. Cancer* 50: 80-85; Posner *et al.*, (1993) *Molecular Pharmacology* 45: 673-683; Rendu *et al.*, (1992) *Biol. Pharmacology* 44(5): 881-888; Sauro and

Thomas, (1993) *Life Sciences* 53: 371-376; Sauro and Thomas, (1993) *J. Pharm. and Experimental Therapeutics* 267(3): 119-1125; Wolbring *et al.*, (1994) *J. Biol. Chem.* 269(36): 22470-22472; and Yoneda *et al.*, (1991) *Cancer Research* 51: 4430-4435; all of which are incorporated herein by reference in their entirety, including any drawings.

Other compounds that could be used as modulators include oxindolinones such as those described in U.S. patent application Serial No. 08/702,232 filed August 23, 1996, incorporated herein by reference in its entirety, including any drawings.

RECOMBINANT DNA TECHNOLOGY:

DNA Constructs Comprising a Kinase Nucleic Acid Molecule and Cells Containing These Constructs:

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complementary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or organism that contains an above-described nucleic acid molecule and thereby is capable of expressing a polypeptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a kinase of the invention may be obtained by the above-described methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a kinase of the invention, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a sequence encoding a kinase of the invention) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a gene sequence encoding a kinase of the invention, or (3) interfere with the ability of the gene sequence of a kinase of the invention to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a gene encoding a kinase of the

-77-

invention, transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene encoding a kinase of the invention (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for kinases of the invention. Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119 and the like; suitable phage or bacteriophage vectors may include λgt10, λgt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, and the like. However, under such conditions, the polypeptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express a kinase of the invention (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link the sequence encoding the kinase of the invention to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage λ, the *bla* promoter of the β-lactamase gene sequence of pBR322, and the *cat* promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage λ (P_L and P_R), the *trp*, *λrecA*, acZ, *λacI*, and *gal*

-78-

promoters of *E. coli*, the α-amylase (Ulmanen *et al.*, *J. Bacteriol*. 162: 176-182,
1985) and the ς-28-specific promoters of *B. subtilis* (Gilman *et al.*, *Gene Sequence* 32:
11-20, 1984), the promoters of the bacteriophages of *Bacillus* (Gryczan, in: The
Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982), and *Streptomyces*
promoters (Ward *et al.*, *Mol. Gen. Genet*. 203: 468-478, 1986). Prokaryotic
promoters are reviewed by Glick (*Ind. Microbiot*. 1: 277-282, 1987), Cenatiempo
(*Biochimie* 68: 505-516, 1986), and Gottesman (*Ann. Rev. Genet*. 18: 415-442,
1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome-
binding site upstream of the gene sequence-encoding sequence. Such ribosome-
binding sites are disclosed, for example, by Gold *et al.* (*Ann. Rev. Microbiol*. 35:
365-404, 1981). The selection of control sequences, expression vectors,
transformation methods, and the like, are dependent on the type of host cell used to
express the gene. As used herein, "cell," "cell line," and "cell culture" may be used
interchangeably and all such designations include progeny. Thus, the words
"transformants" or "transformed cells" include the primary subject cell and cultures
derived therefrom, without regard to the number of transfers. It is also understood
that all progeny may not be precisely identical in DNA content, due to deliberate or
inadvertent mutations. However, as defined, mutant progeny have the same
functionality as that of the originally transformed cell.

Host cells which may be used in the expression systems of the present invention are
not strictly limited, provided that they are suitable for use in the expression of the
kinase polypeptide of interest. Suitable hosts may often include eukaryotic cells.
Preferred eukaryotic hosts include, for example, yeast, fungi, insect cells, mammalian
cells either *in vivo*, or in tissue culture. Mammalian cells which may be useful as
hosts include HeLa cells, cells of fibroblast origin such as VERO or CHO-K1, or cells
of lymphoid origin and their derivatives. Preferred mammalian host cells include
SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 332, which may
provide better capacities for correct post-translational processing.

-79-

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences. Another preferred host is an insect cell, for example the *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can be used (Rubin, *Science* 240: 1453-1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of kinases of the invention in insect cells (Jasny, *Science* 238: 1653, 1987; Miller *et al.*, in: *Genetic Engineering*, Vol. 8, Plenum, Setlow *et al.*, eds., pp. 277-297, 1986).

Any of a series of yeast expression systems can be utilized which incorporate promoter and termination elements from the actively expressed sequences coding for glycolytic enzymes that are produced in large quantities when yeast are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. Yeast provides substantial advantages in that it can also carry out post-translational modifications. A number of recombinant DNA strategies exist utilizing strong promoter sequences and high copy number plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian genes and secretes peptides bearing leader sequences (*i.e.*, pre-peptides). Several possible vector systems are available for the expression of kinases of the invention in a mammalian host.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression

-80-

can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

Expression of kinases of the invention in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer *et al.*, *J. Mol. Appl. Gen.* 1: 273-288, 1982); the TK promoter of Herpes virus (McKnight, *Cell* 31: 355-365, 1982); the SV40 early promoter (Benoist *et al.*, *Nature* (London) 290: 304-31, 1981); and the yeast gal4 gene sequence promoter (Johnston *et al.*, *Proc. Natl. Acad. Sci.* (USA) 79: 6971-6975, 1982; Silver *et al.*, *Proc. Natl. Acad. Sci.* (USA) 81: 5951-5955, 1984).

Translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes a kinase of the invention (or a functional derivative thereof) does not contain any intervening codons which are capable of encoding a methionine (*i.e.*, AUG). The presence of such codons results either in the formation of a fusion protein (if the AUG codon is in the same reading frame as the kinase of the invention coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the kinase of the invention coding sequence).

A nucleic acid molecule encoding a kinase of the invention and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA or RNA molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent expression may occur through the integration of the introduced DNA sequence into the host chromosome.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, e.g., antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama (*Mol. Cell. Biol.* 3: 280-289, 1983).

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColEl, pSC101, pACYC 184, πVX; "Molecular Cloning: A Laboratory Manual," 1989, *supra*). Bacillus plasmids include pC194, pC221, pT127, and the like (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982). Suitable *Streptomyces* plasmids include p1J101 (Kendall *et al., J. Bacteriol.* 169: 4177-4183, 1987), and streptomyces bacteriophages such as φC31 (Chater *et al.*, In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary, pp. 45-54, 1986). *Pseudomonas* plasmids are reviewed by John *et al.* (*Rev. Infect. Dis.* 8: 693-704, 1986), and Izaki (*Jpn. J. Bacteriol.* 33: 729-742, 1978).

Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein *et al.*, Miami Wntr. Symp. 19: 265-274, 1982; Broach, In: "The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470, 1981; Broach, *Ceil* 28: 203-204, 1982; Bollon *et al.*, *J. Clin. Hematol. Oncol.* 10: 39-48, 1980; Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608, 1980).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, *i.e.*, transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene(s) results in the production of a kinase of the invention, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

Transgenic Animals:

A variety of methods are available for the production of transgenic animals associated with this invention. DNA can be injected into the pronucleus of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (*e.g.*, the nucleus of a two-cell embryo) following the initiation of cell division (Brinster *et al.*, *Proc. Nat. Acad. Sci.* USA 82: 4438-4442, 1985). Embryos can be infected with viruses, especially retroviruses, modified to carry inorganic-ion receptor nucleotide sequences of the invention.

Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleotide sequences of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and allowed to come to term. Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan *et al.*, *supra*). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout (*Experientia* 47: 897-905, 1991). Other procedures for introduction of DNA into tissues of animals are described in U.S. Patent No. 4,945,050 (Sanford *et al.*, July 30, 1990).

By way of example only, to prepare a transgenic mouse, female mice are induced to superovulate. Females are placed with males, and the mated females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts. Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female mice are paired with vasectomized males. Recipient females are mated at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice (Hammer *et al.*, *Cell* 63: 1099-1112, 1990).

Methods for the culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection also are well known to those of ordinary skill in the art (Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press, 1987).

In cases involving random gene integration, a clone containing the sequence(s) of the invention is co-transfected with a gene encoding resistance. Alternatively, the gene

-84-

encoding neomycin resistance is physically linked to the sequence(s) of the invention. Transfection and isolation of desired clones are carried out by any one of several methods well known to those of ordinary skill in the art (E.J. Robertson, *supra*).

DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination (Capecchi, *Science* 244: 1288-1292, 1989). Methods for positive selection of the recombination event (*i.e.*, neo resistance) and dual positive-negative selection (*i.e.*, neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Capecchi, *supra* and Joyner *et al.* (*Nature* 338: 153-156, 1989), the teachings of which are incorporated herein in their entirety including any drawings. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others (Houdebine and Chourrout, *supra*; Pursel *et al.*, *Science* 244: 1281-1288, 1989; and Simms *et al.*, *Bio/Technology* 6: 179-183, 1988).

Thus, the invention provides transgenic, nonhuman mammals containing a transgene encoding a kinase of the invention or a gene affecting the expression of the kinase. Such transgenic nonhuman mammals are particularly useful as an *in vivo* test system for studying the effects of introduction of a kinase, or regulating the expression of a kinase (*i.e.*, through the introduction of additional genes, antisense nucleic acids, or ribozymes).

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a cell, which DNA becomes part of the genome of the animal which develops from that cell. Preferred transgenic animals are primates, mice, rats, cows, pigs, horses, goats, sheep, dogs and cats. The transgenic DNA may encode human kinases. Native expression in an animal may be reduced by providing an amount of antisense RNA or DNA effective to reduce expression of the receptor.
Gene Therapy:

-85-

Kinases or their genetic sequences will also be useful in gene therapy (reviewed in Miller, *Nature* 357: 455-460, 1992). Miller states that advances have resulted in practical approaches to human gene therapy that have demonstrated positive initial results. The basic science of gene therapy is described in Mulligan (*Science* 260: 926-931, 1993).

In one preferred embodiment, an expression vector containing a kinase coding sequence is inserted into cells, the cells are grown *in vitro* and then infused in large numbers into patients. In another preferred embodiment, a DNA segment containing a promoter of choice (for example a strong promoter) is transferred into cells containing an endogenous gene encoding kinases of the invention in such a manner that the promoter segment enhances expression of the endogenous kinase gene (for example, the promoter segment is transferred to the cell such that it becomes directly linked to the endogenous kinase gene).

The gene therapy may involve the use of an adenovirus containing kinase cDNA targeted to a tumor, systemic kinase increase by implantation of engineered cells, injection with kinase-encoding virus, or injection of naked kinase DNA into appropriate tissues.

Target cell populations may be modified by introducing altered forms of one or more components of the protein complexes in order to modulate the activity of such complexes. For example, by reducing or inhibiting a complex component activity within target cells, an abnormal signal transduction event(s) leading to a condition may be decreased, inhibited, or reversed. Deletion or missense mutants of a component, that retain the ability to interact with other components of the protein complexes but cannot function in signal transduction, may be used to inhibit an abnormal, deleterious signal transduction event.

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences (*e.g.*, cDNA) encod-ing recombinant kinase of the invention protein into the targeted cell

population (*e.g.*, tumor cells). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing coding sequences (Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y., 1989; Ausubel *et al.*, Current Proto-cols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y., 1989). Alter-natively, recombinant nucleic acid molecules encoding protein sequences can be used as naked DNA or in a recon-stituted system *e.g.*, liposomes or other lipid systems for delivery to target cells (*e.g.*, Felgner *et al.*, *Nature* 337: 387-8, 1989). Several other methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins (Miller, *supra*).

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA into the nucleus of a cell, through a process of microinjection (Capecchi, *Cell* 22: 479-88, 1980). Once recombinant genes are introduced into a cell, they can be recognized by the cell's normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been attempted for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with calcium phosphate and taken into cells by pinocytosis (Chen *et al.*, *Mol. Cell Biol.* 7: 2745-52, 1987); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu *et al.*, *Nucleic Acids Res.* 15: 1311-26, 1987); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner *et al.*, *Proc. Natl. Acad. Sci. USA.* 84: 7413-7417, 1987); and particle bombardment using DNA bound to small projectiles (Yang *et al.*, *Proc. Natl. Acad. Sci.* 87: 9568-9572, 1990). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially

-87-

improves the uptake and expression of the recombinant gene (Curiel *et al.*, *Am. J. Respir. Cell. Mol. Biol.*, 6: 247-52, 1992).

As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, antisense DNA or RNA, or enzymatically active RNA. Gene transfer can be performed in cultured cells or by direct administration into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cyto-plasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer can be performed *ex vivo* on cells which are then transplanted into a patient, or can be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a kinase polypeptide is provided in which the nucleic acid sequence is expressed only in specific tissue. Methods of achieving tissue-specific gene expression are set forth in International Publication No. WO 93/09236, filed November 3, 1992 and published May 13, 1993.

In all of the preceding vectors set forth above, a further aspect of the invention is that the nucleic acid sequence contained in the vector may include additions, deletions or modifications to some or all of the sequence of the nucleic acid, as defined above.

Expression, including over-expression, of a kinase polypeptide of the invention can be inhibited by administration of an antisense molecule that binds to and inhibits expression

-88-

of the mRNA encoding the polypeptide. Alternatively, expression can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. General methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art. Each of these methods utilizes a system, such as a vector, encoding either an antisense or ribozyme transcript of a kinase polypeptide of the invention.

The term "*ribozyme*" refers to an RNA structure of one or more RNAs having catalytic properties. Ribozymes generally exhibit endonuclease, ligase or polymerase activity. Ribozymes are structural RNA molecules which mediate a number of RNA self-cleavage reactions. Various types of trans-acting ribozymes, including "hammerhead" and "hairpin" types, which have different secondary structures, have been identified. A variety of ribozymes have been characterized. See, for example, U.S. Pat. Nos. 5,246,921, 5,225,347, 5,225,337 and 5,149,796. Mixed ribozymes comprising deoxyribo and ribooligonucleotides with catalytic activity have been described. Perreault, et al., Nature, 344: 565-567 (1990).

As used herein, "antisense" refers of nucleic acid molecules or their derivatives which specifically hybridize, e.g., bind, under cellular conditions, with the genomic DNA and/or cellular mRNA encoding a kinase polypeptide of the invention, so as to inhibit expression of that protein, for example, by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix.

In one aspect, the antisense construct is an nucleic acid which is generated ex vivo and that, when introduced into the cell, can inhibit gene expression by, without limitation, hybridizing with the mRNA and/or genomic sequences of a kinase polynucleotide of the invention.

Antisense approaches can involve the design of oligonucleotides (either DNA or RNA) that are complementary to kinase polypeptide mRNA and are based on the kinase polynucleotides of the invention, including SEQ ID NO: 1 through 66. The

-89-

antisense oligonucleotides will bind to the kinase polypeptide mRNA transcripts and prevent translation.

Although absolute complementarity is preferred, it is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

In general, oligonucleotides that are complementary to the 5' end of the message, e.g., the 5' untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3' untranslated sequences of mRNAs have been shown to be effective at inhibiting translation of mRNAs as well. (Wagner, R. (1994) Nature 372: 333). Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5', 3' or coding region of the kinase polypeptide mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably less than about 100 and more preferably less than about 50 or 30 nucleotides in length. Typically they should be between 10 and 25 nucleotides in length. Such principles will inform the practitioner in selecting the appropriate oligonucleotides In preferred embodiments, the antisense sequence is selected from an oligonucleotide sequence that comprises, consists of, or consists essentially of about 10-30, and more preferably 15-25, contiguous nucleotide bases of a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 66 or domains thereof.

In another preferred embodiment, the invention includes an isolated, enriched or purified nucleic acid molecule comprising, consisting of or consisting essentially of about 10-30, and more preferably 15-25 contiguous nucleotide bases of a nucleic acid sequence that encodes a polypeptide of SEQ ID NO: 67 through 132.

Using the sequences of the present invention, antisense oligonucleotides can be designed. Such antisense oligonucleotides would be administered to cells expressing the target kinase and the levels of the target RNA or protein with that of an internal control RNA or protein would be compared. Results obtained using the antisense oligonucleotide would also be compared with those obtained using a suitable control oligonucleotide. A preferred control oligonucleotide is an oligonucleotide of approximately the same length as the test oligonucleotide. Those antisense oligonucleotides resulting in a reduction in levels of target RNA or protein would be selected.

The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci.* U.S.A. 86: 6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci.* USA 84: 648-652; PCT Publication No. WO 88/09810, published Dec. 15, 1988) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134, published Apr. 25, 1988), hybridization-triggered cleavage agents. (*See, e.g.*, Krol *et al.* (1988) BioTechniques 6: 958-976) or intercalating agents. (*See, e.g*, Zon (1988) *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The antisense oligonucleotide may comprise at least one modified base moiety which is selected from moieties such as 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, and 5-(carboxyhydroxyethyl)

uracil. The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof. (*see also* U.S. Pat. Nos. 5,176,996; 5,264,564; and 5,256,775)

In yet a further embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier *et al.* (1987) *Nucl. Acids Res.* 15: 6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue *et al.* (1987) *Nucl. Acids Res.* 15: 6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) FEBS Lett. 215: 327-330).

Also suitable are peptidyl nucleic acids, which are polypeptides such as polyserine, polythreonine, etc. including copolymers containing various amino acids, which are substituted at side-chain positions with nucleic acids (T,A,G,C,U). Chains of such polymers are able to hybridize through complementary bases in the same manner as natural DNA/RNA. Alternatively, an antisense construct of the present invention can be delivered, for example, as an expression plasmid or vector that, when transcribed in the cell, produces RNA complementary to at least a unique portion of the cellular mRNA which encodes a kinase polypeptide of the invention.

While antisense nucleotides complementary to the kinase polypeptide coding region sequence can be used, those complementary to the transcribed untranslated region are most preferred.

In another preferred embodiment, a method of gene replacement is set forth. "Gene replacement" as used herein means supplying a nucleic acid sequence which is capable of being expressed *in vivo* in an animal and thereby providing or augmenting the function of an endogenous gene which is missing or defective in the animal.

PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

The compounds described herein, including kinase polypeptides of the invention, antisense molecules, ribozymes, and any other compound that modulates the activity of a kinase polypeptide of the invention, can be administered to a human patient *per se*, or in pharmaceutical compositions where it is mixed with other active ingredients, as in combination therapy, or suitable carriers or excipient(s). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition.

Routes Of Administration:

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intravenous, intramedullary injections, as well as intrathecal, direct intraventricular, intraperitoneal, intranasal, or intraocular injections.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a solid tumor, often in a depot or sustained release formulation.

Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with tumor-specific antibody. The liposomes will be targeted to and taken up selectively by the tumor.

Composition/Formulation:

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

-93-

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Suitable carriers include excipients such as, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl- cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as

-94-

glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes.

Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD: D5W) consists of VPD diluted 1: 1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the

-96-

fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose.

Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

Many of the tyrosine or serine/threonine kinase modulating compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Pharmaceutically compatible salts may be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms.

Suitable Dosage Regimens:

Pharmaceutical compositions suitable for use in the present invention include compositions where the active ingredients are contained in an amount effective to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount of compound effective to prevent, alleviate or ameliorate symptoms

of disease or prolong the survival of the subject being treated. Determination of a therapeutically effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

Methods of determining the dosages of compounds to be administered to a patient and modes of administering compounds to an organism are disclosed in U.S. Application Serial No. 08/702,282, filed August 23, 1996 and International patent publication number WO 96/22976, published August 1 1996, both of which are incorporated herein by reference in their entirety, including any drawings, figures or tables. Those skilled in the art will appreciate that such descriptions are applicable to the present invention and can be easily adapted to it.

The proper dosage depends on various factors such as the type of disease being treated, the particular composition being used and the size and physiological condition of the patient. Therapeutically effective doses for the compounds described herein can be estimated initially from cell culture and animal models. For example, a dose can be formulated in animal models to achieve a circulating concentration range that initially takes into account the IC_{50} as determined in cell culture assays. The animal model data can be used to more accurately determine useful doses in humans.

For any compound used in the methods of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the tyrosine or serine/threonine kinase activity). Such information can be used to more accurately determine useful doses in humans.

Toxicity and therapeutic efficacy of the compounds described herein can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed

-98-

as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics," Ch. 1 p.1).

In another example, toxicity studies can be carried out by measuring the blood cell composition. For example, toxicity studies can be carried out in a suitable animal model as follows: 1) the compound is administered to mice (an untreated control mouse should also be used); 2) blood samples are periodically obtained via the tail vein from one mouse in each treatment group; and 3) the samples are analyzed for red and white blood cell counts, blood cell composition and the percent of lymphocytes versus polymorphonuclear cells. A comparison of results for each dosing regime with the controls indicates if toxicity is present.

At the termination of each toxicity study, further studies can be carried out by sacrificing the animals (preferably, in accordance with the American Veterinary Medical Association guidelines Report of the American Veterinary Medical Assoc. Panel on Euthanasia: 229-249, 1993). Representative animals from each treatment group can then be examined by gross necropsy for immediate evidence of metastasis, unusual illness or toxicity. Gross abnormalities in tissue are noted and tissues are examined histologically. Compounds causing a reduction in body weight or blood components are less preferred, as are compounds having an adverse effect on major organs. In general, the greater the adverse effect the less preferred the compound.

For the treatment of cancers the expected daily dose of a hydrophobic pharmaceutical agent is between 1 to 500 mg/day, preferably 1 to 250 mg/day, and most preferably 1 to 50 mg/day. Drugs can be delivered less frequently provided plasma levels of the active moiety are sufficient to maintain therapeutic effectiveness.

Plasma levels should reflect the potency of the drug. Generally, the more potent the compound the lower the plasma levels necessary to achieve efficacy.

Plasma half-life and biodistribution of the drug and metabolites in the plasma, tumors and major organs can also be determined to facilitate the selection of drugs most appropriate to inhibit a disorder. Such measurements can be carried out. For example, HPLC analysis can be performed on the plasma of animals treated with the drug and the location of radiolabeled compounds can be determined using detection methods such as X-ray, CAT scan and MRI. Compounds that show potent inhibitory activity in the screening assays, but have poor pharmacokinetic characteristics, can be optimized by altering the chemical structure and retesting. In this regard, compounds displaying good pharmacokinetic characteristics can be used as a model.

Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the kinase modulating effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data; *e.g.*, the concentration necessary to achieve 50-90% inhibition of the kinase using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%.

In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

Packaging:

-100-

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. The pack or dispenser may also be accompanied with a notice associated with the container in form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the polynucleotide for human or veterinary administration. Such notice, for example, may be the labeling approved by the U.S. Food and Drug Administration for prescription drugs, or the approved product insert. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition. Suitable conditions indicated on the label may include treatment of a tumor, inhibition of angiogenesis, treatment of fibrosis, diabetes, and the like.

FUNCTIONAL DERIVATIVES

Also provided herein are functional derivatives of a polypeptide or nucleic acid of the invention. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or binding activity mediated through noncatalytic domains, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide that retains the functionality of the original. In both cases, all permutations are intended to be covered by this disclosure.

Included within the scope of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules. The degeneracy of the genetic code permits substitution of certain codons by other codons that specify the same amino

-101-

acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the genes of the invention could be synthesized to give a nucleic acid sequence significantly different from one selected from the group consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66. The encoded amino acid sequence thereof would, however, be preserved.

In addition, the nucleic acid sequence may comprise a nucleotide sequence which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula selected from the group consisting of those set forth in SEQ ID NO: 1 through SEQ ID NO: 66, or a derivative thereof. Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence of selected from the group consisting of those set forth in SEQ ID NO: 1 through 66, which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the inventive nucleic acid sequence or its derivative, or from the addition of TTA, TAG or TGA as a termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, the nucleic acid molecule of the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end.

Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto. All variations of the nucleotide sequence of the kinase genes of the invention and fragments thereof permitted by the genetic code are, therefore, included in this invention.

Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity as the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two

-102-

polypeptides are functionally equivalent, as are the two nucleic acid molecules that give rise to their production, even though the differences between the nucleic acid molecules are not related to the degeneracy of the genetic code.

A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, as described below.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect or reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin.

-103-

Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine alpha-amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimide (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking the component peptides of the protein to each other or to other proteins in a complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl) dithiolpropioimidate yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates

-104-

described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the proteins, of the complexes having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably, the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. Fragments of a protein are useful for screening for substances that act to modulate signal transduction, as described herein. It is understood that such fragments may retain one or more characterizing portions of the native complex. Examples of such retained characteristics include: catalytic activity; substrate specificity; interaction with other molecules in the intact cell; regulatory functions; or binding with an antibody specific for the native complex, or an epitope thereof.

Another functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant

-105-

may be derived from a naturally occurring complex component by appropriately
modifying the protein DNA coding sequence to add, remove, and/or to modify codons
for one or more amino acids at one or more sites of the C-terminus, N-terminus,
and/or within the native sequence. It is understood that such variants having added,
substituted and/or additional amino acids retain one or more characterizing portions of
the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino
acid residues may be prepared using standard techniques well-known to those of
ordinary skill in the art. For example, the modified components of the functional
derivatives may be produced using site-directed mutagenesis techniques (as
exemplified by Adelman *et al.*, 1983, *DNA* 2: 183) wherein nucleotides in the DNA
coding the sequence are modified such that a modified coding sequence is modified,
and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host
cell, using techniques such as those described above. Alternatively, proteins with
amino acid deletions, insertions and/or substitutions may be conveniently prepared by
direct chemical synthesis, using methods well-known in the art. The functional
derivatives of the proteins typically exhibit the same qualitative biological activity as
the native proteins.

<div align="center">

TABLES
AND
DESCRIPTION THEREOF

</div>

This patent application describes 66 protein kinase polypeptides identified in genomic
and cDNA sequence databases. The results are summarized in six tables, described
below. The Tables appear beginning at page 233.

Table 1 documents the name of each gene, the nucleic acid and amino acid sequence
identification numbers, the species (human or mouse), the classifications of each gene
(superfamily, family and group), the lengths of the nucleic acid and protein
sequences, the positions and lengths of the open reading frames within the sequence,
and whether Sugen has cloned a full length version of the gene. From left to right the
data presented is as follows: Gene name, Species, ID#na, SEQ ID NO: , Super-

<div align="center">

-106-

</div>

family, Group, Family, NA_length, AA_length, ORF Start, ORF End, ORF Length, Physical Status (FL indicates a full-length cDNA version of the gene has been obtained). "Gene name" refers to name given the sequence encoding the kinase or kinase-like enzyme. The "ID#na" and "ID#aa" refer to the SEQ ID NOS given each nucleic acid and amino acid sequence in this patent. "Superfamily" identifies whether the gene is a protein kinase or protein-kinase-like. "Group" and "Family" refer to the protein kinase classification defined by sequence homology and based on previously established phylogenetic analysis [Hardie, G. and Hanks S. The Protein Kinase Book, Academic Press (1995) and Hunter T. and Plowman, G. Trends in Biochemical Sciences (1977) 22: 18-22 and Plowman G.D. et al. (1999) Proc. Natl. Acad. Sci. 96: 13603-13610)]. "NA_length" refers to the length in nucleotides of the corresponding nucleic acid sequence. "AA length" refers to the length in amino acids of the peptide encoded in the corresponding nuclei acid sequence. "ORF start" refers to the beginning nucleotide of the open reading frame. "ORF end" refers to the last nucleotide of the open reading frame, excluding the stop codon. "ORF length" refers to the length in nucleotides of the open reading frame (including the stop codon). In the "Physical Status" column, "FL" indicates a full-length cDNA version of the gene has been obtained.

Table 2 describes the results of Smith Waterman similarity searches (Matrix: Pam100; gap open/extension penalties 12/2) of the amino acid sequences against the NCBI database of non-redundant protein sequences (http: //www.ncbi.nlm.nih.gov/Entrez/protein.html). It is broken into two sections, Tables 2a and 2b. For Table 2a: from left to right the data presented is as follows: Gene_NAME, Species, ID#na, ID#aa, Super-family, Group, Family, AA length, PSCORE, MATCHES, % Identity, % Similarity, ACCESSION, and DESCRIPTION. The first columns (Gene_NAME, Species, ID#na, ID#aa, Super-family, Group, Family, AA length) are the same as in Table 1. "PSCORE" refers to the Smith Waterman probability score. This number approximates the chance that the alignment occurred by chance. Thus, a very low number, such as 2.10E-64, indicates that there is a very significant match between the query and the database target. "Matches" indicates the number of amino acids that were identical in the alignment. "%

Identity" lists the percent of amino acids that were identical over the alignment. "%
Similarity" lists the percent of amino acids that were similar over the alignment.
ACCESSION refers to the accession number of the most similar protein in the NCBI
database of non-redundant proteins. "Description" contains the name and species of
origin of the most similar protein in the NCBI database of non-redundant proteins.
Table 2b continues the tabulation of the Smith Waterman results. The headings are:
Gene_NAME, Species, ID#na, ID#aa, Super-family, Group, Family, QUERYSTART,
QUERYEND, TARGETSTART, TARGETEND, %QUERY, %TARGET. The
"QUERY" is the patent sequence, and the "TARGET" is the best hit within the NCBI
protein database. "QUERYSTART" refers to the amino acid number at which the
Query (the patent protein sequence) begins to align with the TARGET (database)
sequence. "QUERYEND" refers to the amino acid position within the patent protein
sequence (the QUERY) at which the alignment with the database protein (the
TARGET) ends. " TARGETSTART" refers to the amino acid position of the
database protein (the TARGET) at which the alignment with the patent sequence (the
QUERY) begins. "TARGETEND" refers to the amino acid position within the
database sequence (the TARGET) at which alignment with the QUERY ends.
%QUERY gives the percent of the patent amino acid sequence which is aligned with
the database hit (the TARGET). %TARGET gives the percent of the database hit
which aligns with the patent sequence.

Table 3 lists the results of searching the database of single nucleotide polymorphisms
(dbSNP) with the patent nucleic acid sequences. The column headings are: Gene,
ID#na, ID#aa, Nucleotide #, Polymorphism, Nucleotide in patent sequence, AA
Residue #, Silent / Residue Change, AA Residue in Patent, Accession#. "Nucleotide
#" refers the to the position within the nucleic acid sequence at which the SNP occurs;
"Polymorphism" describes the sequence change at the site of the SNP, for example, a
change from C to T; "Nucleotide in patent sequence" lists the nucleotide (A,C,G,T)
present in the patent sequence; "AA Residue #" refers to the position within the patent
protein of the amino acid affected by the SNP (regions outside the coding sequence
are referred to as untranslated regions, or UTRs); "Silent / Residue Change" lists the
nature of the change in the protein sequence as a consequence of the SNP: silent (for

-108-

example "no change," E/A (a glutamic acid in one form is replace by an alanine in the other form), R/stop (a codon for arginine has been altered to a stop codon); "AA Residue in Patent" lists which of the alternative amino acids is present in the patent protein sequence; "Accession#" lists the dbSNP accession number (http: //www.ncbi.nlm.nih.gov/SNP/index.html).

Table 4 describes the extent and the boundaries of the kinase catalytic domains, and other protein domains. These domains were identified using PFAM (http: //pfam.wustl.edu/hmmsearch.shtml) models, a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Version Pfam 7.3 (May 2002) contains alignments and models for 3849 protein families. The PFAM alignments were downloaded from http: //pfam.wustl.edu/hmmsearch.shtml and the HMMr searches were run locally on a Timelogic computer (TimeLogic Corporation, Incline Village, NV). The column headings are: "Gene," "ID#na," " ID#aa," "Profile Description," "Profile Accession," "Pscore," "Domain Start," "Domain End," "Profile Start," "Profile End," "Profile Length," and "Query Length." The "Profile Description" column contains the name of the protein domain; "Profile Accession" refers to the PFAM accession number for the domain; "Pscore" lists the probability score, or E-value, and is the number of hits that would be expected to have a score equal or better by chance alone. A good E-value is much less than 1. Around 1 is what is expected just by chance; "Domain Start" lists the amino acid number within the protein sequence at which the domain begins; "Domain End" lists the amino acid number within the protein sequence at which the domain ends; "Profile Start" refers to the position within the profile at which it begins alignment with the patent sequence; "Profile End" lists the position within the profile at which it the alignment with the patent sequence ends; "Profile Length" lists the length in amino acid residues of the PFAM profile; and "Query Length" lists the amino acid length of the patent protein.

Table 5 lists the chromosomal position of the patent genes. The cytogenetic localization of the kinase genes allows one to compare their map position with databases of "disease loci," such as the "Online Mendelian Inheritance in Man" (http:

-109-

//www.ncbi.nlm.nih.gov/Omim/searchomim.html). This database is a catalog of human genes and genetic disorders maintained at the National Center for Biotechnology Information. The database contains textual information, pictures, and reference information. The column headings for table 5 are: "Gene_Name," "Species," "ID#na," " ID#aa," "Cytogenetic position," "Cancer Amplicon," and "Disease Loci." "Cytogenetic position" lists the cytogenetic band to which the gene has been mapped, "Cancer Amplicon" annotates the observation that the kinase maps to a known cancer amplicon; and "Disease Loci" annotates the observation that the kinase maps to a region implicated in human disease and documented in OMIM.

Table 6 lists human ESTs representing the patent genes. The column headings are: "RANK" (number of ESTs per gene, 1-10 for most; SGK110 and SGK069 were not represented in dbEST database); "Gene" (Gene name and ID numbers); "Human EST" (derived from BLASTN search of http: //www.ncbi.nlm.nih.gov/dbEST/index.html).

EXAMPLES

The examples below are not limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the isolation and characterization of the nucleic acid molecules according to the invention, as well as the polypeptides they encode.

EXAMPLE 1: Identification and Characterization of Genomic Fragments Encoding Protein Kinases

Novel kinases were identified from the Celera human genomic sequence databases, and from the public Human Genome Sequencing project (http: //www.ncbi.nlm.nih.gov/) using a hidden Markov model (HMMR) built with 70 mammalian and yeast kinase catalytic domain sequences. These sequences were chosen from a comprehensive collection of kinases such that no two sequences had more than 50% sequence identity. The genomic database entries were translated in six open reading frames and searched against the model using a Timelogic Decypher box with a Field programmable array (FPGA) accelerated version of HMMR2.1. The DNA sequences encoding the predicted protein sequences aligning to the HMMR profile were extracted from the original genomic database. The nucleic acid sequences were then clustered using the Pangea Clustering tool to eliminated

-110-

repetitive entries. The putative protein kinase sequences were then sequentially run through a series of queries and filters to identify novel protein kinase sequences. Specifically, the HMMR identified sequences were searched using BLASTN and BLASTX against a nucleotide and amino acid repository containing 634 known human protein kinases and all subsequent new protein kinase sequences as they are identified. The output was parsed into a spreadsheet to facilitate elimination of known genes by manual inspection. Two models were developed, a "complete" model and a "partial" or Smith Waterman model. The partial model was used to identify sub-catalytic kinase domains, whereas the complete model was used to identify complete catalytic domains. The selected hits were then queried using BLASTN against the public nrna and EST databases to confirm they are indeed unique. In some cases the novel genes were judged to be homologues of previously identified rodent or vertebrate protein kinases.

Extension of partial DNA sequences to encompass the full-length open-reading frame was carried out by several methods. Iterative blastn searching of the cDNA databases listed in Table 9 was used to find cDNAs that extended the genomic sequences. "LifeSeqGold" databases are from Incyte Genomics, Inc (http: //www.incyte.com/). NCBI databases are from the National Center for Biotechnology Information (http: //www.ncbi.nlm.nih.gov/). All blastn searches were conducted using a penalty for a nucleotide mismatch of –3 and reward for a nucleotide match of 1. The gapped blast algorithm is described in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Res. 25: 3389-3402).

Extension of partial DNA sequences to encompass the full-length open-reading frame was also carried out by iterative searches of genomic databases. The first method made use of the Smith-Waterman algorithm to carry out protein-protein searches of a close protein homologue to the partial. The target databases consisted of Genscan and open-reading frame (ORF) predictions of all human genomic sequence derived from the human genome project (HGP) as well as from Celera. The complete set of

genomic databases searched is shown in Table 7, below. Genomic sequences
encoding potential extensions were further assessed by blastx analysis against the
NCBI nonredundant database to confirm the novelty of the hit. The extending
genomic sequences were incorporated into the cDNA sequence after removal of
potential introns using the Seqman program from DNAStar. The default parameters
used for Smith-Waterman searches were as shown next. Matrix: blosum 62; gap-
opening penalty: 12; gap extension penalty: 2. Genscan predictions were made
using the Genscan program as detailed in Chris Burge and Sam Karlin "Prediction of
Complete Gene Structures in Human Genomic DNA," JMB (1997) 268(1): 78-94).
ORF predictions from genomic DNA were made using a standard 6-frame translation.

Another method for defining DNA extensions from genomic sequence used iterative
searches of genomic databases through the Genscan program to predict exon splicing.
These predicted genes were then assessed to see if they represented "real" extensions
of the partial genes based on homology to related kinases.

Another method involved using the Genewise program (http:
//www.sanger.ac.uk/Software/Wise2/) to predict potential ORFs based on homology
to the closest orthologue/homologue. Genewise requires two inputs, the homologous
protein, and genomic DNA containing the gene of interest. The genomic DNA was
identified by blastn searches of Celera and Human Genome Project databases. The
orthologs were identified by blastp searches of the NCBI non-redundant protein
database (NRAA). Genewise compares the protein sequence to a genomic DNA
sequence, allowing for introns and frameshifting errors.

TABLE 7

Databases used for cDNA-based sequence extensions

Database	Database Date
LifeGold templates	March 2002
LifeGold compseqs	March 2002
LifeGold fl	March 2002
LifeGold flft	March 2002
NCBI human Ests	March 2001
NCBI murine Ests	March 2002
NCBI nonredundant	March 2002

TABLE 8

Databases used for genomic-based sequence extensions

Database	Number of entries	Database Date
Celera Assembly 6	479,986	March 2002
HGP Chromosomal assemblies	2759	March 2002

Results:

For genes that were extended using Genewise, the accession numbers of the protein ortholog and the genomic DNA are given. (Genewise uses the ortholog to assemble the coding sequence of the target gene from the genomic sequence). The amino acid sequences for the orthologs were obtained from the NCBI non-redundant database of proteins .(http: //www.ncbi.nlm.nih.gov/Entrez/protein.html). The genomic DNA came from two sources: Celera andHGP (human genome project), as indicated below. cDNA sources are also listed below. All of the genomic sequences were used

-113-

as input for Genscan predictions to predict splice sites [Burge and Karlin, JMB (1997) 268(1): 78-94)]. Abbreviations: HGP: Human Genome Project; NCBI, National Center for Biotechnology Information.

The results are detailed in the paragraphs below for each gene.

Results – Nucleic Acid Sequences

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the DMPK family. The nucleic acid sequence is 8656 nucleotides long, and codes for a protein that is 2055 amino acids long. The open reading frame starts at nucleotide number 51 and ends at nucleotide number 6218. The length of the ORF is 6168 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 12q24.31. The CRIK sequence maps to Celera contig 181000000794572. A mouse homolog (Rho/rac interacting citron kinase gi|3599509) of CRIK is 353 AAs longer at the N terminus than the public CRIK. Rho/rac interacting citron kinase from mouse (gi|3599509) was used as a model for a genewise prediction. Incyte template, 233643.1, and Incyte CB1 sequence, 7484498CB1, were used to extend the C-terminus of the genewise prediction. Two additional public ESTs (gi|4534019 and gi|3753446) support a different 3' end. These two public ESTs (gi|4534019 and gi|3753446) have an earlier polyA site, just after ATTCTTAATAGATTTGAATAGCGACGTA (just following the run of T's), this generates an alternative 3' end in that form.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the DMPK family. The nucleic acid sequence is 5438 nucleotides long, and codes for a protein that is 1572 amino acids long. The open reading frame starts at nucleotide number 66 and ends at nucleotide number 4784. The length of the ORF is 4719 nucleotides. The gene has been mapped to chromosomal region 11q12-q13.1. This region has been identified as a cancer amplicon (Knuutila, et al). This region has been associated with susceptibility to osteoarthritis (OMIM 165720).

DMPK2 maps to Celera assembly 5 contig 92000004065166. A genewise prediction was run with this contig and myotonic dystrophy associated protein kinase from rat (gi|7446379) as the model. The rat sequence is 118 AA longer at the N-term and 1200 AA longer at the C-term.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the MAST family. The nucleic acid sequence is 5990 nucleotides long, and codes for a protein that is 1332 amino acids long. The open reading frame starts at nucleotide number 36 and ends at nucleotide number 4031. The length of the ORF is 3996 nucleotides. The gene has been mapped to chromosomal region 19p13.1.

The current MAST3 sequence adds a novel N-terminus of 46 AA to sequences previously published. This region is predicted to be of functional importance due to the high level of similarity seen in an orthologous mouse EST (gi|6631994).

MAST205, SEQ ID NO: 4, SEQ ID NO: 70, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the MAST family. The nucleic acid sequence is 5516 nucleotides long, and codes for a protein that is 1798 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 5397. The length of the ORF is 5397 nucleotides. The gene has been mapped to chromosomal region 1p34.1. The public MAST205 sequence is partial at the N and C-terminus. The MAST205 sequence maps to Celera assembly 5 contig 92000004111345. The mouse homolog microtubule-associated testis specific S/T protein kinase (gi|6678958) was used as a model for a genewise prediction.

MASTL, SEQ ID NO: 5, SEQ ID NO: 71, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the MAST family. The nucleic acid sequence is 3882 nucleotides long, and codes for a protein that is 878 amino acids long. The open reading frame starts at nucleotide number 967 and ends at nucleotide number 3603. The length of the ORF is 2637 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to

chromosomal region 10p11.2-p12.1. This region has been associated with susceptibility to schizophrenia (OMIM 181500).

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the PKC family. The nucleic acid sequence is 2392 nucleotides long, and codes for a protein that is 683 amino acids long. The open reading frame starts at nucleotide number 407 and ends at nucleotide number 2458. The length of the ORF is 2052 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 14q23.1.

H19102, SEQ ID NO: 7, SEQ ID NO: 73, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the RSK family. The nucleic acid sequence is 1564 nucleotides long, and codes for a protein that is 449 amino acids long. The open reading frame starts at nucleotide number 188 and ends at nucleotide number 1537. The length of the ORF is 1350 nucleotides. The gene has been mapped to chromosomal region 17q11.1. This region has been identified as a cancer amplicon (Knuutila, et al).

Genewise predictions with the nearest homologs (bicoid-interacting protein in fly and a C. elegans predicted protein) as models yielded some downstream sequence, extending the kinase domain.

MSK1, SEQ ID NO: 8, SEQ ID NO: 74, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the RSK family. The nucleic acid sequence is 3813 nucleotides long, and codes for a protein that is 802 amino acids long. The open reading frame starts at nucleotide number 159 and ends at nucleotide number 2567. The length of the ORF is 2409 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 14q32.11.

YANK3, SEQ ID NO: 9, SEQ ID NO: 75, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the YANK family. The nucleic acid sequence is 2051 nucleotides long, and codes for a protein that is 486

amino acids long. The open reading frame starts at nucleotide number 70 and ends at nucleotide number 1530. The length of the ORF is 1461 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 10q26.3.

MARK2, SEQ ID NO: 10, SEQ ID NO: 76, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the CAMKL family. The nucleic acid sequence is 3063 nucleotides long, and codes for a protein that is 787 amino acids long. The open reading frame starts at nucleotide number 399 and ends at nucleotide number 2762. The length of the ORF is 2364 nucleotides. The gene has been mapped to chromosomal region 11q12-11q13 . This region has been identified as a cancer amplicon (Knuutila, et al). This region has been associated with susceptibility to osteoarthritis (OMIM 165720).

The current sequence extends the N-terminus of published sequences by 33 AA. The mouse ortholog (gi|6679643) is identical in these 33 AA, which implies that this terminal region is important for full biological function of the protein and has been highly conserved to preserve that function.

NuaK2, SEQ ID NO: 11, SEQ ID NO: 77, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the CAMKL family. The nucleic acid sequence is 3463 nucleotides long, and codes for a protein that is 672 amino acids long. The open reading frame starts at nucleotide number 57 and ends at nucleotide number 2075. The length of the ORF is 2019 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 1q31-q32.1.

BRSK2, SEQ ID NO: 12, SEQ ID NO: 78, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the CAMKL family. The nucleic acid sequence is 3831 nucleotides long, and codes for a protein that is 674 amino acids long. The open reading frame starts at nucleotide number 25 and ends at nucleotide number 2049. The length of the ORF is 2025 nucleotides. The gene has been mapped to chromosomal region 11p15.5.

MARK4, SEQ ID NO: 13, SEQ ID NO: 79, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the CAMKL family. The nucleic acid sequence is 3249 nucleotides long, and codes for a protein that is 752 amino acids long. The open reading frame starts at nucleotide number 17 and ends at nucleotide number 2275. The length of the ORF is 2259 nucleotides. The gene has been mapped to chromosomal region 19q13.2-q13.33. This region has been identified as a cancer amplicon (Knuutila, et al).

DCAMKL2, SEQ ID NO: 14, SEQ ID NO: 80, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the DCAMKL family. The nucleic acid sequence is 2827 nucleotides long, and codes for a protein that is 766 amino acids long. The open reading frame starts at nucleotide number 350 and ends at nucleotide number 2650. The length of the ORF is 2301 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 4q31.3.

PIM2, SEQ ID NO: 15, SEQ ID NO: 81, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the PIM family. The nucleic acid sequence is 2186 nucleotides long, and codes for a protein that is 435 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 1305. The length of the ORF is 1305 nucleotides. The gene has been mapped to chromosomal region Xp11.23. This region has been identified as a cancer amplicon (Knuutila, et al).

Based on other family members, and rodent orthologs it has been determined that the PIM2 protein starts with an atypical CTG initiation codon, making the first AA an L rather than an M.

PIM3, SEQ ID NO: 16, SEQ ID NO: 82, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the PIM family. The nucleic acid sequence is 2405 nucleotides long, and codes for a protein that is 326 amino acids long. The open reading frame starts at nucleotide number 436 and ends at nucleotide number 1416. The length of the ORF is 981 nucleotides. Sugen has

-118-

cloned the full length cDNA for this gene. The gene has been mapped to chromosomal region 22q13.

TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, is a member of the Protein Kinase superfamily. It is further classified into the CAMK group, and the TSSK family. The nucleic acid sequence is 1710 nucleotides long, and codes for a protein that is 328 amino acids long. The open reading frame starts at nucleotide number 617 and ends at nucleotide number 1603. The length of the ORF is 987 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 14q11.1.

The ORF was also extended by documenting an alternative splice variant (7693857.2) which shortened the 5' end of exon 4 by 72 nucleotides(splicing out an inframe stop codon): >72 alternatively spliced nucleotides GTCCAACTGCTCATTGCCTGTGTGGCACAATGGAGAAAAACTCAGGCAAG ACCTCTCTCTCCCCTGCTCTAG. Canonical splice sites are maintained with both splice variants. The sequence now shares tight similarity to a mouse cDNA from RIKEN (gi|12855865) over its full length.

CKIL2, SEQ ID NO: 18, SEQ ID NO: 84, is a member of the Protein Kinase superfamily. It is further classified into the CKI group, and the CKIL family. The nucleic acid sequence is 5946 nucleotides long, and codes for a protein that is 1244 amino acids long. The open reading frame starts at nucleotide number 368 and ends at nucleotide number 4102. The length of the ORF is 3735 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 15q14-q15.3. This region has been associated with susceptibility to schizophrenia (OMIM 181500). \

PCTAIRE3, SEQ ID NO: 19, SEQ ID NO: 85, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the CDK family. The nucleic acid sequence is 3229 nucleotides long, and codes for a protein that is 505 amino acids long. The open reading frame starts at nucleotide number 303 and ends at nucleotide number 1817. The length of the ORF is 1515 nucleotides. The full

length cDNA for this gene has been cloned.   The gene has been mapped to chromosomal region 1q32.

PFTAIRE2, SEQ ID NO: 20, SEQ ID NO: 86, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the CDK family. The nucleic acid sequence is 2250 nucleotides long, and codes for a protein that is 435 amino acids long. The open reading frame starts at nucleotide number 45 and ends at nucleotide number 1352. The length of the ORF is 1308 nucleotides. The full length cDNA for this gene has been cloned.   The gene has been mapped to chromosomal region 2q33.2-q34. This region has been identified as a cancer amplicon (Knuutila, et al). This region has been associated with susceptibility to osteoarthritis (OMIM 140600).

ERK7, SEQ ID NO: 21, SEQ ID NO: 87, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the MAPK family. The nucleic acid sequence is 1906 nucleotides long, and codes for a protein that is 563 amino acids long. The open reading frame starts at nucleotide number 19 and ends at nucleotide number 1710. The length of the ORF is 1692 nucleotides. The full length cDNA for this gene has been cloned.   The gene has been mapped to chromosomal region 8q24.3. A genewise prediction was run with a rat homolog, extracellular signal-regulated kinase 7 (gi|4220888), as the model. Two splice variants were noted for ERK7: >Nucleotides 967 – 1098 are alternatively spliced GCACTGCAGCACCCCTACGTGCAGAGGTTCCACTGCCCCAGCGACGAGTG GGCACGAGAGGCAGATGTGCGGCCCCGGGCACACGAAGGGGTCCAGCTC TCTGTGCCTGAGTACCGCAGCCGCGTCTATCAG. >Nucleotides 184 – 240 are alternatively spliced GACATGGGCTTCCTTCTTGCTCCACCCACCCACACACCTGTGTTTCTGTCTC TTCAG.

CKIIa-rs, SEQ ID NO: 22, SEQ ID NO: 88, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the CKII family. The nucleic acid sequence is 1494 nucleotides long, and codes for a protein that is 391 amino acids long. The open reading frame starts at nucleotide number 150 and ends

at nucleotide number 1325. The length of the ORF is 1176 nucleotides. The gene has been mapped to chromosomal region 11p15.

DYRK4, SEQ ID NO: 23, SEQ ID NO: 89, is a member of the Protein Kinase superfamily. It is further classified into the CMCG group, and the DYRK family. The nucleic acid sequence is 2886 nucleotides long, and codes for a protein that is 921 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 2766. The length of the ORF is 2766 nucleotides. The full length cDNA for this gene was cloned. The gene has been mapped to chromosomal region 12p13. This region has been associated with susceptibility to essential hypertension (OMIM 145500).

HIPK1, SEQ ID NO: 24, SEQ ID NO: 90, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the DYRK family. The nucleic acid sequence is 8212 nucleotides long, and codes for a protein that is 1210 amino acids long. The open reading frame starts at nucleotide number 286 and ends at nucleotide number 3918. The length of the ORF is 3633 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 1p11-p12. Contigs from Celera and HGP with homeoedomain interacting protein kinase 1 from mouse were used for genewise predictions.

HIPK4, SEQ ID NO: 25, SEQ ID NO: 91, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the DYRK family. The nucleic acid sequence is 3142 nucleotides long, and codes for a protein that is 616 amino acids long. The open reading frame starts at nucleotide number 977 and ends at nucleotide number 2827. The length of the ORF is 1851 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 19q13.1. This region has been identified as a cancer amplicon (Knuutila, et al).

BIKE, SEQ ID NO: 26, SEQ ID NO: 92, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NAK family. The nucleic acid sequence is 3895 nucleotides long, and codes for a protein that is 1161

amino acids long. The open reading frame starts at nucleotide number 203 and ends at nucleotide number 3688. The length of the ORF is 3486 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 4q13-q21.21. This region has been associated with susceptibility to osteoarthritis (OMIM 140600).

The BIKE sequence is full length, and 89% identical to murine BIKE across the full length of the protein.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NEK family. The nucleic acid sequence is 3912 nucleotides long, and codes for a protein that is 1125 amino acids long. The open reading frame starts at nucleotide number 176 and ends at nucleotide number 3553. The length of the ORF is 3378 nucleotides. The full length cDNA for this gene has been cloned.   The gene has been mapped to chromosomal region 3p21.33.

pNEK5, SEQ ID NO: 28, SEQ ID NO: 94, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NEK family. The nucleic acid sequence is 2816 nucleotides long, and codes for a protein that is 889 amino acids long. The open reading frame starts at nucleotide number 147 and ends at nucleotide number 2816. The length of the ORF is 2670 nucleotides. The full length cDNA for this gene has been cloned.   The gene has been mapped to chromosomal region 13q14. This region has been identified as a cancer amplicon (Knuutila, et al).

The current sequence is an extension of our previously filed patent application sequence (gi|14546899, Sequence 45 from Patent WO0138503), incorporated herein by reference, which adds a 57 AA extension to the N terminus, a 127 AA extension to the C-terminus and is alternatively spliced at two regions in the middle of the gene.

NEK1, SEQ ID NO: 29, SEQ ID NO: 95, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NEK family. The nucleic acid sequence is 5583 nucleotides long, and codes for a protein that is 1286

-122-

amino acids long. The open reading frame starts at nucleotide number 493 and ends at nucleotide number 4353. The length of the ORF is 3861 nucleotides. The gene has been mapped to chromosomal region 4q33-q34.

The revised sequence now contains a complete kinase domain and overlaps completely with the mouse ortholog of Nek1 (gi|1709251). Three alternative splice variants were noted: >Nucleotides 243 – 320 (canonical splice sites maintained) gtgtggagagtctcagtgcccccctttcagtctggactgtgagctgctgctggttagacagtcttggtttctctttcag. >Nucleotides 1923 – 2054 (canonical splice sites maintained) AGGAATTCTGCCTGGAGTTCGTCCAGGATTTCCTTATGGGGCTGCAGGTCA TCACCATTTTCCTGATGCTGATGATATTAGAAAAACTTTGAAAAGATTGAA GGCGGTGTCTAAACAAGCCAATGCAAACAG. >Nucleotides 2158 – 2241 (canonical splice sites maintained). GGAATCCTGCAAAACCTGGCAGCTATGTATGGAGGCAGGCCCAGCTCTTC AAGAGGAGGGAAGCCAAGAAACAAAGAGGAAGAG.

NEK3, SEQ ID NO: 30, SEQ ID NO: 96, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NEK family. The nucleic acid sequence is 2326 nucleotides long, and codes for a protein that is 506 amino acids long. The open reading frame starts at nucleotide number 296 and ends at nucleotide number 1816. The length of the ORF is 1521 nucleotides. The gene has been mapped to chromosomal region 13q14.3. This region has been identified as a cancer amplicon (Knuutila, et al).

SGK069, SEQ ID NO: 31, SEQ ID NO: 97, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NKF1 family. The nucleic acid sequence is 1156 nucleotides long, and codes for a protein that is 348 amino acids long. The open reading frame starts at nucleotide number 110 and ends at nucleotide number 1156. The length of the ORF is 1047 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 19q13.43.

SGK110, SEQ ID NO: 32, SEQ ID NO: 98, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NKF1 family. The nucleic acid sequence is 1853 nucleotides long, and codes for a protein that is 414 amino acids long. The open reading frame starts at nucleotide number 299 and ends at nucleotide number 1543. The length of the ORF is 1245 nucleotides. Sugen has cloned the full length cDNA for this gene. The gene has been mapped to chromosomal region 19q13.43.

NRBP2, SEQ ID NO: 33, SEQ ID NO: 99, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NRBP family. The nucleic acid sequence is 3765 nucleotides long, and codes for a protein that is 507 amino acids long. The open reading frame starts at nucleotide number 282 and ends at nucleotide number 1805. The length of the ORF is 1524 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 8q24.3.

CNK, SEQ ID NO: 34, SEQ ID NO: 100, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the PLK family. The nucleic acid sequence is 2535 nucleotides long, and codes for a protein that is 646 amino acids long. The open reading frame starts at nucleotide number 534 and ends at nucleotide number 2474. The length of the ORF is 1941 nucleotides. The gene has been mapped to chromosomal region 1p34.1.

Two alternative splice variants were noted (Incyte template 222139.15): (1) an intron read through over the intron between exons 9 and 10, (2) exon 6 is alternatively spliced:

>Nucleotides (insert after nucleotide 1697)
GTGAGGCGCTCAGGTGGACACTGTTCCCCTGACTCACCCCCACCCTAGCA
GCTGAGGGAAGCCGGGGATAAAAGAGGCTGCTGAAGCATCCAGCCTCGT
GGTGGCCTAATTGGCTGTGTGTCACCAGCCTGGCGGGGCTGACCTGGGGT
GCCCTGGGAGCCAGGGCAGGGCCAGGCCATGGACTCAAGGGTTTGGATTT
TGGGGCCTGTGTCACTCCCTTTCCCTGCCCAACCCTCCAG

-124-

>Nucleotides 2039 –2168

GACTGTGCACTACAATCCCACCAGCACAAAGCACTTCTCCTTCTCCGTGGG
TGCTGTGCCCCGGGCCCTGCAGCCTCAGCTGGGTATCCTGCGGTACTTCGC
CTCCTACATGGAGCAGCACCTCATGAAG

SCYL2, SEQ ID NO: 35, SEQ ID NO: 101, is a member of the Protein Kinase
superfamily. It is further classified into the Other group, and the SCY1 family. The
nucleic acid sequence is 5525 nucleotides long, and codes for a protein that is 933
amino acids long. The open reading frame starts at nucleotide number 173 and ends
at nucleotide number 2974. The length of the ORF is 2802 nucleotides. The gene has
been mapped to chromosomal region 12q23-q24.1.

SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, is a member of the Protein Kinase
superfamily. It is further classified into the CMGC group, and the SRPK family. The
nucleic acid sequence is 3715 nucleotides long, and codes for a protein that is 688
amino acids long. The open reading frame starts at nucleotide number 179 and ends
at nucleotide number 2245. The length of the ORF is 2067 nucleotides. The full
length cDNA for this gene has been cloned.   The gene has been mapped to
chromosomal region 7q22.3. This region has been identified as a cancer amplicon
(Knuutila, et al).

TLK1, SEQ ID NO: 37, SEQ ID NO: 103, is a member of the Protein Kinase
superfamily. It is further classified into the Other group, and the TLK family. The
nucleic acid sequence is 4321 nucleotides long, and codes for a protein that is 787
amino acids long. The open reading frame starts at nucleotide number 238 and ends
at nucleotide number 2601. The length of the ORF is 2364 nucleotides.   The gene
has been mapped to chromosomal region 2q31.1. This region has been associated
with susceptibility to osteoarthritis (OMIM 140600).

One alternative splice variant was noted:
>Nucleotides 645 – 707

GTTCCCCAACCTCCCGGTCTTCCAGTCCTTGGCCTATTGGGAAATGGGTCG
TACAGCAGGAGG.

SGK071, SEQ ID NO: 38, SEQ ID NO: 104, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Unique family. The nucleic acid sequence is 2285 nucleotides long, and codes for a protein that is 632 amino acids long. The open reading frame starts at nucleotide number 195 and ends at nucleotide number 2093. The length of the ORF is 1899 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 9q34.

SK516, SEQ ID NO: 39, SEQ ID NO: 105, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Unique family. The nucleic acid sequence is 7364 nucleotides long, and codes for a protein that is 929 amino acids long. The open reading frame starts at nucleotide number 180 and ends at nucleotide number 2969. The length of the ORF is 2790 nucleotides. The gene has been mapped to chromosomal region 1q31-32.1.

H85389, SEQ ID NO: 40, SEQ ID NO: 106, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the ULK family. The nucleic acid sequence is 1971 nucleotides long, and codes for a protein that is 401 amino acids long. The open reading frame starts at nucleotide number 134 and ends at nucleotide number 1339. The length of the ORF is 1206 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 20p13.

Wee1b, SEQ ID NO: 41, SEQ ID NO: 107, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the WEE family. The nucleic acid sequence is 1704 nucleotides long, and codes for a protein that is 567 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 1704. The length of the ORF is 1704 nucleotides. The gene has been mapped to chromosomal region 7q34-36.

Wnk2, SEQ ID NO: 42, SEQ ID NO: 108, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Wnk family. The nucleic acid sequence is 7981 nucleotides long, and codes for a protein that is 2245

amino acids long. The open reading frame starts at nucleotide number 67 and ends at nucleotide number 6804. The length of the ORF is 6738 nucleotides. The gene has been mapped to chromosomal region 9q22.31. Other members of this family (Wnk1 and Wnk4) have been strongly implicated in hypertension (Lifton RP, et al, Human hypertension caused by mutations in WNK kinases, Science. 2001 Aug 10;293(5532): 1107-12), and so Wnk2 may also play a role in this disease.

Six alternative splice variants are noted:

> Wnk2, SEQ ID NO: 42 Nucleotides 2059 and 2214

CCTGGCTTGCCGGTGGGCTCTGTCCCGGCCCCCGCCTGCCCTCCGTCCCTC
CAGCAGCACTTCCCGGATCCGGCCATGAGCTTCGCCCCCGTGCTGCCGCC
GCCCAGCACCCCCATGCCCACGGGCCCAGGCCAGCCAGCACCCCCGGCC
AGCAG

>Wnk2, SEQ ID NO: 42Nucleotides 5945 and 6136

GTCACTTGGCTGACTCCAGCAGAGGCCCTCCCGCTAAGGACCCTGCCCAA
GCCAGTGTGGGGCTCACTGCAGACAGCACGGGCCTGAGCGGGAAGGCAG
TGCAGACCCAGCAGCCCTGCTCCGTCCGGGCCTCCCTGTCTTCGGACATCT
GCTCCGGCTTAGCCAGTGATGGAGGCGGAGCGCGTGGCCAAG

> Wnk2, SEQ ID NO: 42Nucleotides 6137 and 6280

GCTGGACGGTTTACCACCCAACGTCTGAGAGAGTGACCTATAAGTCTAGT
AGCAAACCTCGTGCTCGATTCCTCAGTGGACCCGTATCTGTGTCCATCTGG
TCTGCCCTGAAGCGTCTCTGCCTAGGCAAAGAACACAGCAGTA

> Wnk2, SEQ ID NO: 42 Nucleotides 5945 and 6280

GTCACTTGGCTGACTCCAGCAGAGGCCCTCCCGCTAAGGACCCTGCCCAA
GCCAGTGTGGGGCTCACTGCAGACAGCACGGGCCTGAGCGGGAAGGCAG
TGCAGACCCAGCAGCCCTGCTCCGTCCGGGCCTCCCTGTCTTCGGACATCT
GCTCCGGCTTAGCCAGTGATGGAGGCGGAGCGCGTGGCCAAGGCTGGACG
GTTTACCACCCAACGTCTGAGAGAGTGACCTATAAGTCTAGTAGCAAACC

TCGTGCTCGATTCCTCAGTGGACCCGTATCTGTGTCCATCTGGTCTGCCCT
GAAGCGTCTCTGCCTAGGCAAAGAACACAGCAGTA

> Wnk2, SEQ ID NO: 42 Insert after nucleotide 620
TCTGTGCGGTTGACTCCTTTTCCTCCCCGCCTGGAGATCCCCGTGGTGTCG
ACTGGAAGCATGGAGGCACCTTGGGGAG

> Wnk2, SEQ ID NO: 42 Replaces nucleotides 6650 – 7981
ATCCTGAGAGTGAGAAGCCTGACTGACCCCGCCTAGACGCCAGGCCCACT
TCACGCCGTCTAAGTGGAGAAGTGACGGACCCTCAGGGCCAGCTGCTCCT
CCTGTCCAGTTCACGCTGTTTTGTAACCACTTTCTAAGCATTTTTTATTCAC
AATTGGAAACACAAATGTAATGCAAGAATAAAAAATATTTTGGGGCAGA
AAGGACTTTGGTTTTTCAAACTATTTCCTCTCTGGTGGCCCTCGGCCAGCC
AGGTGACTGGGATGTGACAGGGGTGGGGGGACATTCCCAGGACCCTGGC
ATGCTCAGGATAGCCCTGTTCTCTGCAGGGCCCTGGAGGTGGCGGCCCCG
GGGAGGCTGATCTCCAAGTCCCCCCGATGCCAGCTGGC

MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, is a member of the Protein Kinase
superfamily. It is further classified into the STE group, and the STE11 family. The
nucleic acid sequence is 7026 nucleotides long, and codes for a protein that is 1511
amino acids long. The open reading frame starts at nucleotide number 1 and ends at
nucleotide number 4536. The length of the ORF is 4536 nucleotides. The gene has
been mapped to chromosomal region 5q11.2-q13. This region has been associated
with susceptibility to schizophrenia (OMIM 181500).

The sequence has good similarity to the mouse and rat orthologs.

MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, is a member of the Protein Kinase
superfamily. It is further classified into the STE group, and the STE11 family. The
nucleic acid sequence is 2571 nucleotides long, and codes for a protein that is 735
amino acids long. The open reading frame starts at nucleotide number 1 and ends at
nucleotide number 2208. The length of the ORF is 2208 nucleotides. The gene has
been mapped to chromosomal region 2q21.3.

One alternative splice variant was noted:

>MAP3K8, SEQ ID NO: 44 Replaces nucleotides 1412 – 2571

GTTCAAGTCCAATGGGAAAGAAATATCTTCCTTCAACAGCTGAATATGTT
ACTGGAAGTTTGGAGAATCATTACTAGATGGCAAAAACAAAAGATGTTCC
TTCCATTTTGTGAACTGCATAAGAGATCTTGGGGGGTGGGCGATGAAGAG
AGGTATACTGTGGTCTCACTAGTCAAGGACAGCTAATAGCTGTAAAACAG
GTGGCTTTGGATAACT

Pak4_m, SEQ ID NO: 45 SEQ ID NO: 111, is the only murine sequence in this application. It is a member of the Protein Kinase superfamily, further classified into the STE group, and the STE20 family. The nucleic acid sequence is 1782 nucleotides long, and codes for a protein that is 593 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 1782. The length of the ORF is 1782 nucleotides. The human ortholog has been mapped to 19q13.2.

STLK6-rs, SEQ ID NO: 46 SEQ ID NO: 112, is a member of the Protein Kinase superfamily. It is further classified into the STE group, and the STE20 family. The nucleic acid sequence is 2171 nucleotides long, and codes for a protein that is 418 amino acids long. The open reading frame starts at nucleotide number 242 and ends at nucleotide number 1498. The length of the ORF is 1257 nucleotides. The gene has been mapped to chromosomal region 1p33.

MAP2K2, SEQ ID NO: 47 SEQ ID NO: 113, is a member of the Protein Kinase superfamily. It is further classified into the STE group, and the STE7 family. The nucleic acid sequence is 1724 nucleotides long, and codes for a protein that is 380 amino acids long. The open reading frame starts at nucleotide number 248 and ends at nucleotide number 1390. The length of the ORF is 1143 nucleotides. Sugen has cloned the full length cDNA for this gene. The gene has been mapped to chromosomal region 7q34.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, is a member of the Protein Kinase superfamily. It is further classified into the TK group, and the CCK4 family. The nucleic acid sequence is 4232 nucleotides long, and codes for a protein that is 1070

amino acids long. The open reading frame starts at nucleotide number 191 and ends at nucleotide number 3403. The length of the ORF is 3213 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 6p21-p12 .

LMR1, SEQ ID NO: 49 SEQ ID NO: 115, is a member of the Protein Kinase superfamily. It is further classified into the TK group, and the Lmr family. The nucleic acid sequence is 5313 nucleotides long, and codes for a protein that is 1374 amino acids long. The open reading frame starts at nucleotide number 85 and ends at nucleotide number 4209. The length of the ORF is 4125 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 17q25.

RYK, SEQ ID NO: 50 SEQ ID NO: 116, is a member of the Protein Kinase superfamily. It is further classified into the TK group, and the Ryk family. The nucleic acid sequence is 3663 nucleotides long, and codes for a protein that is 607 amino acids long. The open reading frame starts at nucleotide number 91 and ends at nucleotide number 1914. The length of the ORF is 1824 nucleotides. The gene has been mapped to chromosomal region 3q22.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, is a member of the Protein Kinase superfamily. It is further classified into the TKL group, and the LRRK family. The nucleic acid sequence is 9753 nucleotides long, and codes for a protein that is 2534 amino acids long. The open reading frame starts at nucleotide number 633 and ends at nucleotide number 8237. The length of the ORF is 7605 nucleotides. The gene has been mapped to chromosomal region 12q11-q12.

For LRRK2, the 3' most 4 nucleotides of the original SGK040 sequence were mispredicted. Correcting the prediction removes the stop and allows for further 3' extension. The sequence was extended at the 3' end by three EST/cDNA sequences (Incyte templates 215217.7 and 215217.9 and NCBI_nr cDNA gi|17454342). Two different splice variants were present. Because the Incyte template 215217.7 and the NCBI_nr cDNA gi|17454342 3' extension yields a longer ORF it was used in the final

-130-

sequence, extending the sequence in the 3' direction by 133 AA and through the stop codon. The 5' most 52 nucleotides of the original sequence were mispredicted and removed from the final revised sequence. The 5' end of the sequence was extended by an overlapping Incyte flft CB1 sequence (71059650CB1) which is supported in two different stretches by over lapping Incyte templates (1017699.1, 316571.1, 415310.1 and 295385.1). Parts of the 5' extension are based on the Incyte CB1 sequence and a genscan prediction. The N-terminus was extended by approximately 1500 AA.

pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, is a member of the Protein Kinase superfamily. It is further classified into the TKL group, and the MLK family. The nucleic acid sequence is 4667 nucleotides long, and codes for a protein that is 1036 amino acids long. The open reading frame starts at nucleotide number 262 and ends at nucleotide number 3372. The length of the ORF is 3111 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 1q42.2.

KSR, SEQ ID NO: 53 SEQ ID NO: 119, is a member of the Protein Kinase superfamily. It is further classified into the TKL group, and the RAF family. The nucleic acid sequence is 5913 nucleotides long, and codes for a protein that is 901 amino acids long. The open reading frame starts at nucleotide number 165 and ends at nucleotide number 2870. The length of the ORF is 2706 nucleotides. The gene has been mapped to chromosomal region 17q11.1. This region has been identified as a cancer amplicon (Knuutila, et al).

The patent sequence for KSR, SEQ ID NO: 53 SEQ ID NO: 119 is full length, and aligns across the full length with the mouse ortholog.

KSR2, SEQ ID NO: 54 SEQ ID NO: 120, is a member of the Protein Kinase superfamily. It is further classified into the TKL group, and the RAF family. The nucleic acid sequence is 2994 nucleotides long, and codes for a protein that is 982 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 2949. The length of the ORF is 2949 nucleotides. The full length

cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 12q24.3.

KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, is a member of the Lipid Kinase superfamily. It is further classified into the DAG kin group, and the DAG kin family. The nucleic acid sequence is 4429 nucleotides long, and codes for a protein that is 537 amino acids long. The open reading frame starts at nucleotide number 92 and ends at nucleotide number 1705. The length of the ORF is 1614 nucleotides. The gene has been mapped to chromosomal region 22q13.31.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, is a member of the Lipid Kinase superfamily. It is further classified into the DAG kin group, and the DAG kin family. The nucleic acid sequence is 4297 nucleotides long, and codes for a protein that is 804 amino acids long. The open reading frame starts at nucleotide number 372 and ends at nucleotide number 2786. The length of the ORF is 2415 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 7p21.3-p22. This region has been associated with susceptibility to osteoarthritis (OMIM 140600).

IP6K1, SEQ ID NO: 57 SEQ ID NO: 123, is a member of the Lipid Kinase superfamily. It is further classified into the Inositol kinase group, and the IP6K family. The nucleic acid sequence is 4461 nucleotides long, and codes for a protein that is 441 amino acids long. The open reading frame starts at nucleotide number 309 and ends at nucleotide number 1634. The length of the ORF is 1326 nucleotides. The gene has been mapped to chromosomal region 3p21.31.

YAB1, SEQ ID NO: 58 SEQ ID NO: 124, is a member of the Atypical PK superfamily. It is further classified into the Atypical group, and the ABC1 family. The nucleic acid sequence is 2508 nucleotides long, and codes for a protein that is 647 amino acids long. The open reading frame starts at nucleotide number 99 and ends at nucleotide number 2042. The length of the ORF is 1944 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal

-132-

region 1q42. This region has been associated with susceptibility to schizophrenia (OMIM 181500).

AF052122, SEQ ID NO: 59 SEQ ID NO: 125, is a member of the Atypical PK superfamily. It is further classified into the Atypical group, and the ABC1 family. The nucleic acid sequence is 5237 nucleotides long, and codes for a protein that is 591 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 1776. The length of the ORF is 1776 nucleotides. Sugen has cloned the full length cDNA for this gene. The gene has been mapped to chromosomal region 19q13.1. This region has been identified as a cancer amplicon (Knuutila, et al).

AAF23326, SEQ ID NO: 60 SEQ ID NO: 126, is a member of the Atypical PK superfamily. It is further classified into the Atypical group, and the ABC1 family. The nucleic acid sequence is 1368 nucleotides long, and codes for a protein that is 455 amino acids long. The open reading frame starts at nucleotide number 1 and ends at nucleotide number 1368. The length of the ORF is 1368 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 14q24.3-q32.

SGK493, SEQ ID NO: 61 SEQ ID NO: 127, is a member of the Atypical PK superfamily. It is further classified into the Atypical group, and the RIO1 family. The nucleic acid sequence is 1832 nucleotides long, and codes for a protein that is 552 amino acids long. The open reading frame starts at nucleotide number 50 and ends at nucleotide number 1708. The length of the ORF is 1659 nucleotides. The full length cDNA for this gene has been cloned. The gene has been mapped to chromosomal region 5q14.

BRD2, SEQ ID NO: 62 SEQ ID NO: 128, is a member of the Atypical PK superfamily. It is further classified into the BRD group, and the BRD family. The nucleic acid sequence is 4693 nucleotides long, and codes for a protein that is 801 amino acids long. The open reading frame starts at nucleotide number 1702 and ends

at nucleotide number 4107. The length of the ORF is 2406 nucleotides. The gene has been mapped to chromosomal region 6p21.2.

BRD3, SEQ ID NO: 63, SEQ ID NO: 129, is a member of the Atypical PK superfamily. It is further classified into the BRD group, and the BRD family. The nucleic acid sequence is 3085 nucleotides long, and codes for a protein that is 726 amino acids long. The open reading frame starts at nucleotide number 140 and ends at nucleotide number 2320. The length of the ORF is 2181 nucleotides. The gene has been mapped to chromosomal region 9q34.

BRD4, SEQ ID NO: 64, SEQ ID NO: 130, is a member of the Atypical PK superfamily. It is further classified into the BRD group, and the BRD family. The nucleic acid sequence is 3149 nucleotides long, and codes for a protein that is 722 amino acids long. The open reading frame starts at nucleotide number 223 and ends at nucleotide number 2391. The length of the ORF is 2169 nucleotides. The gene has been mapped to chromosomal region 19p13.2.

BRDT, SEQ ID NO: 65, SEQ ID NO: 131, is a member of the Atypical PK superfamily. It is further classified into the BRD group, and the BRD family. The nucleic acid sequence is 3106 nucleotides long, and codes for a protein that is 947 amino acids long. The open reading frame starts at nucleotide number 108 and ends at nucleotide number 2951. The length of the ORF is 2844 nucleotides. The gene has been mapped to chromosomal region 1p21.

ZC1, SEQ ID NO: 66, SEQ ID NO: 132 is a member of the protein kinase superfamily, the STE group, and the STE20 family. The nucleic acid sequence is 7986 nucleotides long, and codes for a protein (in its longest form) of 1392 amino acids (see below for splice variants). The open reading frame starts at nucleotide number 366 and ends at nucleotide number 4544. The length of the ORF is 4179 nucleotides. The gene has been mapped to chromosomal region 2q11.1-q11.2.

PolyA tails are present in ZC1, SEQ ID NO: 66 after position 4791, position 6100 and position 7986. All sites are within the 3 prime untranslated region and do not alter the protein sequence. Differential use of these polyadenylation sites has been seen in

ESTs from brain and other tissues, indicating that sequences within the untranslated region may be involved in controlling gene expression in a tissue-specific manner. Alternatively spliced transcripts have been seen in cDNA and EST sequences which lack portions of this sequence. Nine sections (modules) of this sequence are alternatively spliced and it is predicted that transcripts containing all combinations of alternatively spliced modules exist. All alternatively spliced modules are within the open reading frame and contain a multiple of three nucleotides. Therefore, omission of any one module from a transcript results in an inframe deletion of a peptide from the protein. No frameshifts or premature stops are produced by any of these alternatively spliced forms. The positions of the modules on the DNA and protein sequences are as follows:

Module	DNA range	Protein Range	Notes for ZC1, SEQ ID NO: 66
M1	1761-1847	466-494	Encodes C-terminal extension of coiled-coil domain. Similar module found in the paralogous gene TNIK.
M2	1848-1940	495-525	
M3	2070-2231	569-622	Similar module found in TNIK. Contains 2 PxxP motifs, predicted to bind SH3-domain proteins
M4	2232-2462	623-694	Contains 2 PxxP motifs.
M5	2568-2570	736-736	
M6	2821-2829	819-821	
M7	3126-3317	921-984	
M8	4008-4064	1215-1233	Encodes part of CNH domain. Similar sequence seen in other human GCK-IV kinases
M9	4137-4160	1258-1265	Encodes part of CNH domain. Similar sequence not seen in other CNH domains.

EXAMPLE 2a: Expression Analysis of Polypeptides of the Invention

The gene expression patterns for selected genes were studied using a PCR screen of 96 human tissues. This technique does not yield quantitative expression levels between tissues, but does identify which tissues express the gene at a level detectable by PCR and those which do not.

Example 2b: Predicted proteins

II. Predicted Proteins

Description of the Proteins – Smith-Waterman Comparisons (Table 2, a & b)

CRIK, SEQ ID NO: 1, SEQ ID NO: 67 encodes a protein that is 2055 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1975; percent identity over the alignment: 96 %; percent similarity over the alignment: 98 %; accession number for best hit:

-136-

AAC72823.1; description and species for best hit: Rho/rac-interacting citron kinase [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 2055; target start: 1; target end: 2055. The percent of the query that aligns with the target is: 96%. The percent of the target that aligns with the query is: 96%.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68 encodes a protein that is 1572 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 2.20E-211; number of matches: 731; percent identity over the alignment: 45 % ; percent similarity over the alignment: 63 % ; accession number for best hit: NP_446109.1; description and species for best hit: Ser-Thr protein kinase related to the myotonic dystrophy protein kinase [Rattus norvegicus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 2; query end: 1462; target start: 4; target end: 1588. The percent of the query that aligns with the target is: 46%. The percent of the target that aligns with the query is: 42%.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69 encodes a protein that is 1331 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1287; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: BAA25487.1; description and species for best hit: (AB011133) KIAA0561 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 39; query end: 1331; target start: 16; target end: 1308. The percent of the query that aligns with the target is: 96%. The percent of the target that aligns with the query is: 98%.

MAST205, SEQ ID NO: 4, SEQ ID NO: 70 encodes a protein that is 1798 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1684; percent identity over the alignment:

-137-

99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: NP_055927.1; description and species for best hit: KIAA0807 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1687; target start: 1; target end: 1687. The percent of the query that aligns with the target is: 93%. The percent of the target that aligns with the query is: 97%.

MASTL, SEQ ID NO: 5, SEQ ID NO: 71 encodes a protein that is 878 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 876; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: NP_116233.1; description and species for best hit: Hypothetical protein FLJ14813 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 878; target start: 1; target end: 878. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72 encodes a protein that is 683 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 679; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: NP_006246.1; description and species for best hit: (NM_006255) protein kinase C, eta [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 683; target start: 1; target end: 682. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

H19102, SEQ ID NO: 7, SEQ ID NO: 73 encodes a protein that is 449 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.00E-124; number of matches: 269; percent identity over the alignment: 99

-138-

% ; percent similarity over the alignment: 99 % ; accession number for best hit: BAB71555.1; description and species for best hit: Unnamed protein product [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 41; query end: 310; target start: 1; target end: 271. The percent of the query that aligns with the target is: 59%. The percent of the target that aligns with the query is: 98%.

MSK1, SEQ ID NO: 8, SEQ ID NO: 74 encodes a protein that is 802 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 3.50E-304; number of matches: 787; percent identity over the alignment: 98 % ; percent similarity over the alignment: 98 % ; accession number for best hit: NP_004746.1; description and species for best hit: (NM_004755) ribosomal protein S6 kinase, 90kD, polypeptide 5; mitogen- and stress-activated protein kinase 1 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 800; target start: 1; target end: 800. The percent of the query that aligns with the target is: 98%. The percent of the target that aligns with the query is: 97%.

YANK3, SEQ ID NO: 9, SEQ ID NO: 75 encodes a protein that is 486 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 8.9e-311 ; number of matches: 444; percent identity over the alignment: 91 % ; percent similarity over the alignment: 94 % ; accession number for best hit: AAH26457; description and species for best hit: (BC026457) hypothetical serine/threonine protein kinase [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 485; target start: 1; target end: 487. The percent of the query that aligns with the target is: 91%. The percent of the target that aligns with the query is: 90%.

MARK2, SEQ ID NO: 10, SEQ ID NO: 76 encodes a protein that is 787 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein

database with the amino acid sequence for this protein yielded the following results: P score = 2.60E-299; number of matches: 752; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: AAH08771.1; description and species for best hit: (BC008771) Similar to ELKL motif kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 34; query end: 787; target start: 1; target end: 755. The percent of the query that aligns with the target is: 95%. The percent of the target that aligns with the query is: 99%.

NuaK2, SEQ ID NO: 11, SEQ ID NO: 77 encodes a protein that is 672 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 5.10E-269; number of matches: 628; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_112214.1; description and species for best hit: (NM_030952) hypothetical protein DKFZp434J037 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 45; query end: 672; target start: 1; target end: 628. The percent of the query that aligns with the target is: 93%. The percent of the target that aligns with the query is: 100%.

BRSK2, SEQ ID NO: 12, SEQ ID NO: 78 encodes a protein that is 674 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 4.20E-175; number of matches: 602; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: CAA07196.1; description and species for best hit: Putative serine/threonine protein kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 72; query end: 674; target start: 1; target end: 603. The percent of the query that aligns with the target is: 89%. The percent of the target that aligns with the query is: 99%.

MARK4, SEQ ID NO: 13, SEQ ID NO: 79 encodes a protein that is 752 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 4.30E-298; number of matches: 751; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: AAL23683.1; description and species for best hit: MARK4 serine/threonine protein kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 752; target start: 1; target end: 752. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

DCAMKL2, SEQ ID NO: 14, SEQ ID NO: 80 encodes a protein that is 766 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 8.10E-159; number of matches: 513; percent identity over the alignment: 67 % ; percent similarity over the alignment: 80 % ; accession number for best hit: O15075; description and species for best hit: DCAMKL1 (doublecortin-like and CAMK-like 1) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 741; target start: 1; target end: 739. The percent of the query that aligns with the target is: 66%. The percent of the target that aligns with the query is: 69%.

PIM2, SEQ ID NO: 15, SEQ ID NO: 81 encodes a protein that is 434 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.40E-145; number of matches: 334; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_006866.1; description and species for best hit: (NM_006875) pim-2 oncogene; proto-oncogene Pim-2 (serine threonine kinase) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 101; query end: 434; target start: 1; target end: 334. The

percent of the query that aligns with the target is: 76%. The percent of the target that aligns with the query is: 100%.

PIM3, SEQ ID NO: 16, SEQ ID NO: 82 encodes a protein that is 326 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 9.90E-174; number of matches: 311; percent identity over the alignment: 95 % ; percent similarity over the alignment: 97 % ; accession number for best hit: AAH17621.1; description and species for best hit: Serine threonine kinase pim3 [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 326; target start: 1; target end: 326. The percent of the query that aligns with the target is: 95%. The percent of the target that aligns with the query is: 95%.

TSSK4, SEQ ID NO: 17, SEQ ID NO: 83 encodes a protein that is 328 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.60E-69; number of matches: 281; percent identity over the alignment: 85 % ; percent similarity over the alignment: 94 % ; accession number for best hit: BAB30483.1; description and species for best hit: Putative [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 328; target start: 1; target end: 328. The percent of the query that aligns with the target is: 85%. The percent of the target that aligns with the query is: 85%.

CKIL2, SEQ ID NO: 18, SEQ ID NO: 84 encodes a protein that is 1244 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.50E-298; number of matches: 645; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: BAA74870.1; description and species for best hit: KIAA0847 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 600; query end: 1244; target

start: 1; target end: 645. The percent of the query that aligns with the target is: 51%. The percent of the target that aligns with the query is: 100%.

PCTAIRE3, SEQ ID NO: 19, SEQ ID NO: 85 encodes a protein that is 504 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.50E-220; number of matches: 471; percent identity over the alignment: 93 %; percent similarity over the alignment: 93 %; accession number for best hit: Q07002; description and species for best hit: Serine/threonine protein kinase PCTAIRE-3 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 502; target start: 1; target end: 472. The percent of the query that aligns with the target is: 93%. The percent of the target that aligns with the query is: 99%.

PFTAIRE2, SEQ ID NO: 20, SEQ ID NO: 86 encodes a protein that is 435 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 8.40E-100; number of matches: 225; percent identity over the alignment: 68 %; percent similarity over the alignment: 81 %; accession number for best hit: NP_035204.1; description and species for best hit: (NM_011074) PFTAIRE protein kinase 1 [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 97; query end: 426; target start: 129; target end: 458. The percent of the query that aligns with the target is: 51%. The percent of the target that aligns with the query is: 47%.

ERK7, SEQ ID NO: 21, SEQ ID NO: 87 encodes a protein that is 563 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.90E-128; number of matches: 384; percent identity over the alignment: 67 %; percent similarity over the alignment: 75 %; accession number for best hit: AAD12719.2; description and species for best hit: Extracellular signal-regulated kinase 7; ERK7 [Rattus norvegicus]. The boundaries of the alignments for the query

and the database (target) amino acid sequences were as follows. Query start: 1; query end: 560; target start: 1; target end: 544. The percent of the query that aligns with the target is: 68%. The percent of the target that aligns with the query is: 70%.

CKIIa-rs, SEQ ID NO: 22, SEQ ID NO: 88 encodes a protein that is 391 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 9.60E-195; number of matches: 390; percent identity over the alignment: 99 % ; percent similarity over the alignment: 100 % ; accession number for best hit: CAA49758.1; description and species for best hit: Casein kinase II alpha subunit [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 391; target start: 1; target end: 391. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

DYRK4, SEQ ID NO: 23, SEQ ID NO: 89 encodes a protein that is 921 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.20E-304; number of matches: 526; percent identity over the alignment: 99 % ; percent similarity over the alignment: 100 % ; accession number for best hit: Q9NR20; description and species for best hit: DYRK4 4 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 395; query end: 921; target start: 15; target end: 541. The percent of the query that aligns with the target is: 57%. The percent of the target that aligns with the query is: 97%.

HIPK1, SEQ ID NO: 24, SEQ ID NO: 90 encodes a protein that is 1210 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1181; percent identity over the alignment: 97 % ; percent similarity over the alignment: 99 % ; accession number for best hit: AAD41592.1; description and species for best hit: Myak-L [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid

-144-

sequences were as follows. Query start: 1; query end: 1210; target start: 1; target end: 1210. The percent of the query that aligns with the target is: 97%. The percent of the target that aligns with the query is: 97%.

HIPK4, SEQ ID NO: 25, SEQ ID NO: 91 encodes a protein that is 616 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 598; percent identity over the alignment: 97 % ; percent similarity over the alignment: 98 % ; accession number for best hit: BAB72080.1; description and species for best hit: Hypothetical protein [Macaca fascicularis]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 616; target start: 1; target end: 616. The percent of the query that aligns with the target is: 97%. The percent of the target that aligns with the query is: 97%.

BIKE, SEQ ID NO: 26, SEQ ID NO: 92 encodes a protein that is 1161 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 7.60E-244; number of matches: 960; percent identity over the alignment: 82 % ; percent similarity over the alignment: 89 % ; accession number for best hit: NP_542439.1; description and species for best hit: (NM_080708) Bmp2-inducible kinase [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1161; target start: 1; target end: 1138. The percent of the query that aligns with the target is: 82%. The percent of the target that aligns with the query is: 84%.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93 encodes a protein that is 1125 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 9.80E-185; number of matches: 428; percent identity over the alignment: 90 % ; percent similarity over the alignment: 90 % ; accession number for best hit: BAB71395.1; description and species for best hit: (AK057247) unnamed protein product [Homo sapiens]. The boundaries of the alignments for the query and the

database (target) amino acid sequences were as follows. Query start: 698; query end: 1125; target start: 10; target end: 484. The percent of the query that aligns with the target is: 38%. The percent of the target that aligns with the query is: 88%.

pNEK5, SEQ ID NO: 28, SEQ ID NO: 94 encodes a protein that is 889 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.60E-78; number of matches: 180; percent identity over the alignment: 65 % ; percent similarity over the alignment: 82 % ; accession number for best hit: P51954; description and species for best hit: Serine/threonine-protein kinase NEK1 (NimA-related protein kinase 1) [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 58; query end: 333; target start: 1; target end: 275. The percent of the query that aligns with the target is: 20%. The percent of the target that aligns with the query is: 23%.

NEK1, SEQ ID NO: 29, SEQ ID NO: 95 encodes a protein that is 1286 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1258; percent identity over the alignment: 97 % ; percent similarity over the alignment: 97 % ; accession number for best hit: BAB67794.1; description and species for best hit: KIAA1901 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1286; target start: 8; target end: 1265. The percent of the query that aligns with the target is: 97%. The percent of the target that aligns with the query is: 99%.

NEK3, SEQ ID NO: 30, SEQ ID NO: 96 encodes a protein that is 506 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.80E-202; number of matches: 458; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: P51956; description and species for best hit: SERINE/THREONINE-PROTEIN

KINASE NEK3 (NIMA-RELATED PROTEIN KINASE 3) (HSPK 36) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 48; query end: 506; target start: 1; target end: 459. The percent of the query that aligns with the target is: 90%. The percent of the target that aligns with the query is: 99%.

SGK069, SEQ ID NO: 31, SEQ ID NO: 97 encodes a protein that is 348 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 7.40E-48; number of matches: 122; percent identity over the alignment: 42 % ; percent similarity over the alignment: 59 % ; accession number for best hit: AAK52420.1; description and species for best hit: Protein kinase Bsk146 [Danio rerio]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 348; target start: 394; target end: 763. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 41%.

SGK110, SEQ ID NO: 32, SEQ ID NO: 98 encodes a protein that is 414 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 4.00E-35; number of matches: 110; percent identity over the alignment: 41 % ; percent similarity over the alignment: 60 % ; accession number for best hit: S71887; description and species for best hit: serine/threonine-specific kinase (EC 2.7.1.-), pk9.7 gastrula-specific [Xenopus laevis]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 96; query end: 359; target start: 9; target end: 272. The percent of the query that aligns with the target is: 26%. The percent of the target that aligns with the query is: 30%.

NRBP2, SEQ ID NO: 33, SEQ ID NO: 99 encodes a protein that is 507 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 3.20E-158; number of matches: 300; percent identity over the alignment: 61

-147-

% ; percent similarity over the alignment: 75 % ; accession number for best hit: NP_037524.1; description and species for best hit: Nuclear receptor binding protein; multiple domain putative nuclear protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 17; query end: 502; target start: 44; target end: 518. The percent of the query that aligns with the target is: 59%. The percent of the target that aligns with the query is: 56%.

CNK, SEQ ID NO: 34, SEQ ID NO: 100 encodes a protein that is 646 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 8.60E-236; number of matches: 645; percent identity over the alignment: 99 % ; percent similarity over the alignment: 100 % ; accession number for best hit: AAH13899.1; description and species for best hit: (BC013899) Unknown (protein for MGC: 14852) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 646; target start: 1; target end: 646. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

SCYL2, SEQ ID NO: 35, SEQ ID NO: 101 encodes a protein that is 933 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 791; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: BAA92598.1; description and species for best hit: KIAA1360 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 140; query end: 933; target start: 3; target end: 796. The percent of the query that aligns with the target is: 84%. The percent of the target that aligns with the query is: 99%.

SRPK2, SEQ ID NO: 36, SEQ ID NO: 102 encodes a protein that is 688 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P

score = 7.80E-183; number of matches: 684; percent identity over the alignment: 99
% ; percent similarity over the alignment: 99 % ; accession number for best hit:
NP_003129.1; description and species for best hit: (NM_003138) SFRS protein
kinase 2 [Homo sapiens]. The boundaries of the alignments for the query and the
database (target) amino acid sequences were as follows. Query start: 1; query end:
688; target start: 1; target end: 686. The percent of the query that aligns with the
target is: 99%. The percent of the target that aligns with the query is: 99%.

TLK1, SEQ ID NO: 37, SEQ ID NO: 103 encodes a protein that is 787 amino acids
long. The results of a Smith-Waterman search of the NCBI non-redundant protein
database with the amino acid sequence for this protein yielded the following results: P
score = 0; number of matches: 777; percent identity over the alignment: 98 % ;
percent similarity over the alignment: 99 % ; accession number for best hit:
NP_036422.1; description and species for best hit: (NM_012290) tousled-like kinase
1; KIAA0137 gene product; serine threonine protein kinase [Homo sapiens]. The
boundaries of the alignments for the query and the database (target) amino acid
sequences were as follows. Query start: 1; query end: 787; target start: 1; target
end: 787. The percent of the query that aligns with the target is: 98%. The percent
of the target that aligns with the query is: 98%.

SGK071, SEQ ID NO: 38, SEQ ID NO: 104 encodes a protein that is 632 amino
acids long. The results of a Smith-Waterman search of the NCBI non-redundant
protein database with the amino acid sequence for this protein yielded the following
results: P score = 0.000001; number of matches: 63; percent identity over the
alignment: 30 % ; percent similarity over the alignment: 50 % ; accession number
for best hit: NP_175853.1; description and species for best hit: Hypothetical protein
[Arabidopsis thaliana]. The boundaries of the alignments for the query and the
database (target) amino acid sequences were as follows. Query start: 25; query end:
228; target start: 1; target end: 197. The percent of the query that aligns with the
target is: 9%. The percent of the target that aligns with the query is: 10%.

SK516, SEQ ID NO: 39, SEQ ID NO: 105 encodes a protein that is 929 amino acids
long. The results of a Smith-Waterman search of the NCBI non-redundant protein

database with the amino acid sequence for this protein yielded the following results: P score = 5.70E-180; number of matches: 365; percent identity over the alignment: 100 %; percent similarity over the alignment: 100 %; accession number for best hit: BAA32317.1; description and species for best hit: KIAA0472 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 565; query end: 929; target start: 1; target end: 365. The percent of the query that aligns with the target is: 39%. The percent of the target that aligns with the query is: 100%.

H85389, SEQ ID NO: 40, SEQ ID NO: 106 encodes a protein that is 401 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 2.40E-162; number of matches: 400; percent identity over the alignment: 99 %; percent similarity over the alignment: 99 %; accession number for best hit: CAC10518.2; description and species for best hit: Novel protein kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 401; target start: 118; target end: 517. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 77%.

Wee1b, SEQ ID NO: 41, SEQ ID NO: 107 encodes a protein that is 567 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 2.00E-287; number of matches: 541; percent identity over the alignment: 96 %; percent similarity over the alignment: 96 %; accession number for best hit: AAD04726.1; description and species for best hit: Similar to wee1-like protein kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 559; target start: 1; target end: 541. The percent of the query that aligns with the target is: 95%. The percent of the target that aligns with the query is: 100%.

Wnk2, SEQ ID NO: 42, SEQ ID NO: 108 encodes a protein that is 2245 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein

-150-

database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1385; percent identity over the alignment: 99 %; percent similarity over the alignment: 99 %; accession number for best hit: BAB21851.1; description and species for best hit: KIAA1760 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 860; query end: 2245; target start: 1; target end: 1386. The percent of the query that aligns with the target is: 61%. The percent of the target that aligns with the query is: 99%.

MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109 encodes a protein that is 1511 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1459; percent identity over the alignment: 97 %; percent similarity over the alignment: 97 %; accession number for best hit: Q13233; description and species for best hit: MEKK 1 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 21; query end: 1511; target start: 2; target end: 1495. The percent of the query that aligns with the target is: 96%. The percent of the target that aligns with the query is: 97%.

MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110 encodes a protein that is 735 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 2.80E-82; number of matches: 168; percent identity over the alignment: 100 %; percent similarity over the alignment: 100 %; accession number for best hit: XP_017343.1; description and species for best hit: Hypothetical protein fragment FLJ23074 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 547; query end: 714; target start: 1; target end: 168. The percent of the query that aligns with the target is: 22%. The percent of the target that aligns with the query is: 100%.

Pak5_m, SEQ ID NO: 45 SEQ ID NO: 111 encodes a protein that is 593 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant

-151-

protein database with the amino acid sequence for this protein yielded the following results: P score = 2.70E-130; number of matches: 550; percent identity over the alignment: 92 % ; percent similarity over the alignment: 96 % ; accession number for best hit: NP_005875.1; description and species for best hit: p21-activated kinase 4; protein kinase related to S. cerevisiae STE20, effector for Cdc42Hs [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 593; target start: 1; target end: 591. The percent of the query that aligns with the target is: 92%. The percent of the target that aligns with the query is: 93%.

STLK6-rs, SEQ ID NO: 46 SEQ ID NO: 112 encodes a protein that is 418 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 5.90E-222; number of matches: 407; percent identity over the alignment: 97 % ; percent similarity over the alignment: 98 % ; accession number for best hit: NP_061041.2; description and species for best hit: Amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 418; target start: 1; target end: 418. The percent of the query that aligns with the target is: 97%. The percent of the target that aligns with the query is: 97%.

MAP2K2, SEQ ID NO: 47 SEQ ID NO: 113 encodes a protein that is 381 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 4.80E-156; number of matches: 353; percent identity over the alignment: 92 % ; percent similarity over the alignment: 95 % ; accession number for best hit: NP_109587.1; description and species for best hit: (NM_030662) mitogen-activated protein kinase kinase 2; protein kinase, mitogen-activated, kinase 2, p45 (MAP kinase kinase 2) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 2;

query end: 380; target start: 1; target end: 380. The percent of the query that aligns with the target is: 92%. The percent of the target that aligns with the query is: 88%.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114 encodes a protein that is 1070 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1069; percent identity over the alignment: 99 % ; percent similarity over the alignment: 100 % ; accession number for best hit: JC4593; description and species for best hit: protein-tyrosine kinase-related receptor PTK7 precursor [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1070; target start: 1; target end: 1070. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

LMR1, SEQ ID NO: 49 SEQ ID NO: 115 encodes a protein that is 1374 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1207; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_004911.1; description and species for best hit: (NM_004920) apoptosis-associated tyrosine kinase [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 168; query end: 1374; target start: 1; target end: 1207. The percent of the query that aligns with the target is: 87%. The percent of the target that aligns with the query is: 100%.

RYK, SEQ ID NO: 50 SEQ ID NO: 116 encodes a protein that is 607 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 3.60E-287; number of matches: 603; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: I37560; description and species for best hit: Protein-tyrosine kinase Ryk -[Homo sapiens]. The boundaries of the alignments for the query and the database (target)

amino acid sequences were as follows. Query start: 1; query end: 607; target start: 1; target end: 607. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117 encodes a protein that is 2534 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 7.90E-189; number of matches: 463; percent identity over the alignment: 84 % ; percent similarity over the alignment: 92 % ; accession number for best hit: NP_080006.1; description and species for best hit: RIKEN cDNA 4921513O20 gene [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1990; query end: 2534; target start: 17; target end: 561. The percent of the query that aligns with the target is: 18%. The percent of the target that aligns with the query is: 82%.

pMLK4, SEQ ID NO: 52 SEQ ID NO: 118 encodes a protein that is 1036 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1027; percent identity over the alignment: 99 % ; percent similarity over the alignment: 99 % ; accession number for best hit: CAC84640.1; description and species for best hit: (AJ311798) mixed lineage kinase 4 beta [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1036; target start: 1; target end: 1036. The percent of the query that aligns with the target is: 99%. The percent of the target that aligns with the query is: 99%.

KSR, SEQ ID NO: 53 SEQ ID NO: 119 encodes a protein that is 901 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 3.30E-269; number of matches: 797; percent identity over the alignment: 88 % ; percent similarity over the alignment: 92 % ; accession number for best hit: NP_038599.1; description and species for best hit: (NM_013571) kinase suppressor

of ras [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 901; target start: 1; target end: 873. The percent of the query that aligns with the target is: 88%. The percent of the target that aligns with the query is: 91%.

KSR2, SEQ ID NO: 54 SEQ ID NO: 120 encodes a protein that is 982 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 9.60E-119; number of matches: 452; percent identity over the alignment: 48 % ; percent similarity over the alignment: 62 % ; accession number for best hit: NP_038599.1; description and species for best hit: (NM_013571) kinase suppressor of ras [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 51; query end: 982; target start: 34; target end: 849. The percent of the query that aligns with the target is: 46%. The percent of the target that aligns with the query is: 51%.

KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121 encodes a protein that is 537 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 481; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: BAB33316.1; description and species for best hit: KIAA1646 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 57; query end: 537; target start: 1; target end: 481. The percent of the query that aligns with the target is: 89%. The percent of the target that aligns with the query is: 100%.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122 encodes a protein that is 804 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 804; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: Q9Y6T7; description and species for best hit: Diacylglycerol kinase, bets (DGK-

BETA) [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 804; target start: 1; target end: 804. The percent of the query that aligns with the target is: 100%. The percent of the target that aligns with the query is: 100%.

IP6K1, SEQ ID NO: 57 SEQ ID NO: 123 encodes a protein that is 441 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.60E-257; number of matches: 441; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: BAA13393.2; description and species for best hit: KIAA0263 protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 441; target start: 22; target end: 462. The percent of the query that aligns with the target is: 100%. The percent of the target that aligns with the query is: 95%.

YAB1, SEQ ID NO: 58 SEQ ID NO: 124 encodes a protein that is 647 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 3.80E-244; number of matches: 368; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_064632.1; description and species for best hit: (NM_020247) chaperone, ABC1 activity of bc1 complex like [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 280; query end: 647; target start: 1; target end: 368. The percent of the query that aligns with the target is: 56%. The percent of the target that aligns with the query is: 100%.

AF052122, SEQ ID NO: 59 SEQ ID NO: 125 encodes a protein that is 591 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.20E-246; number of matches: 385; percent identity over the alignment: 99 % ; percent similarity over the alignment: 100 % ; accession number

for best hit: AAH13114.1; description and species for best hit: Hypothetical protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 206; query end: 591; target start: 1; target end: 386. The percent of the query that aligns with the target is: 65%. The percent of the target that aligns with the query is: 99%.

AAF23326, SEQ ID NO: 60 SEQ ID NO: 126 encodes a protein that is 455 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 1.40E-304; number of matches: 455; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_065154.1; description and species for best hit: Hypothetical protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 455; target start: 1; target end: 455. The percent of the query that aligns with the target is: 100%. The percent of the target that aligns with the query is: 100%.

SGK493, SEQ ID NO: 61 SEQ ID NO: 127 encodes a protein that is 552 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 552; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit: NP_060813.1; description and species for best hit: Hypothetical protein FLJ11159 [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 552; target start: 1; target end: 552. The percent of the query that aligns with the target is: 100%. The percent of the target that aligns with the query is: 100%.

BRD2, SEQ ID NO: 62 SEQ ID NO: 128 encodes a protein that is 801 amino acids long. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 2.60E-256; number of matches: 801; percent identity over the alignment: 100 % ; percent similarity over the alignment: 100 % ; accession number for best hit:

NP_005095.1; description and species for best hit: Bromodomain-containing protein
2; female sterile homeotic-related gene 1 [Homo sapiens]. The boundaries of the
alignments for the query and the database (target) amino acid sequences were as
follows. Query start: 1; query end: 801; target start: 1; target end: 801. The percent
of the query that aligns with the target is: 100%. The percent of the target that aligns
with the query is: 100%.

BRD3, SEQ ID NO: 63, SEQ ID NO: 129 encodes a protein that is 726 amino acids
long. The results of a Smith-Waterman search of the NCBI non-redundant protein
database with the amino acid sequence for this protein yielded the following results: P
score = 2.20E-243; number of matches: 726; percent identity over the alignment:
100 % ; percent similarity over the alignment: 100 % ; accession number for best hit:
NP_031397.1; description and species for best hit: Bromodomain-containing protein
3 [Homo sapiens]. The boundaries of the alignments for the query and the database
(target) amino acid sequences were as follows. Query start: 1; query end: 726; target
start: 1; target end: 726. The percent of the query that aligns with the target is:
100%. The percent of the target that aligns with the query is: 100%.

BRD4, SEQ ID NO: 64, SEQ ID NO: 130 encodes a protein that is 722 amino acids
long. The results of a Smith-Waterman search of the NCBI non-redundant protein
database with the amino acid sequence for this protein yielded the following results: P
score = 2.60E-232; number of matches: 722; percent identity over the alignment:
100 % ; percent similarity over the alignment: 100 % ; accession number for best hit:
NP_055114.1; description and species for best hit: Bromodomain-containing protein
4 [Homo sapiens]. The boundaries of the alignments for the query and the database
(target) amino acid sequences were as follows. Query start: 1; query end: 722; target
start: 1; target end: 722. The percent of the query that aligns with the target is:
100%. The percent of the target that aligns with the query is: 100%.

BRDT, SEQ ID NO: 65, SEQ ID NO: 131 encodes a protein that is 947 amino acids
long. The results of a Smith-Waterman search of the NCBI non-redundant protein
database with the amino acid sequence for this protein yielded the following results: P
score = 0; number of matches: 947; percent identity over the alignment: 100 % ;

-158-

percent similarity over the alignment: 100 % ; accession number for best hit: NP_001717.1; description and species for best hit: Testis-specific bromodomain protein [Homo sapiens]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 947; target start: 1; target end: 947. The percent of the query that aligns with the target is: 100%. The percent of the target that aligns with the query is: 100%.

ZC1, SEQ ID NO: 66, SEQ ID NO: 132 encodes a protein that is 1392 amino acids long. It has multiple splice variants, as described above in the Nucleic Acids description section. The results of a Smith-Waterman search of the NCBI non-redundant protein database with the amino acid sequence for this protein yielded the following results: P score = 0; number of matches: 1202; percent identity over the alignment: 86 % ; percent similarity over the alignment: 87 % ; accession number for best hit: NP_032722; description and species for best hit: NCK interacting kinase; HPK/GCK-like kinase [Mus musculus]. The boundaries of the alignments for the query and the database (target) amino acid sequences were as follows. Query start: 1; query end: 1392; target start: 1; target end: 12433. The percent of the query that aligns with the target is: 87%. The percent of the target that aligns with the query is: 98%.

Domains of predicted proteins (Table 4)

Many protein kinases contain modular domains in addition to the protein kinases domain. These extra-catalytic domains may play key roles in regulating the activity, protein-protein interactions, and sub-cellular localization of the protein. The paragraphs below describe in detail the protein domains found within the patent sequences. These domains were identified using PFAM (http: //pfam.wustl.edu/hmmsearch.shtml) models, a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Version Pfam 7.3 (May 2002) contains alignments and models for 3849 protein families. The PFAM alignments were downloaded from http: //pfam.wustl.edu/hmmsearch.shtml and the HMMr searches were run locally on a Timelogic computer (TimeLogic Corporation, Incline Village, NV).

Results:

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 9.20E-67. The domain starts at amino acid 98 and ends at amino acid 361. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, has a CNH domain, (PFAM profile accession # PF00780), identified with P_score 2.60E-115. The domain starts at amino acid 1620 and ends at amino acid 1917. The profile has a length of 378 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 378.

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, has a PH domain, (PFAM profile accession # PF00169), identified with P_score 3.00E-16. The domain starts at amino acid 1472 and ends at amino acid 1591. The profile has a length of 85 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 85.

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 1.00E-09. The domain starts at amino acid 1391 and ends at amino acid 1439. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, has a Protein kinase C terminal domain, (PFAM profile accession # PF00433), identified with P_score 3.00E-08. The domain starts at amino acid 362 and ends at amino acid 391. The profile has a length of 70 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 32.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.10E-70. The domain starts at amino acid 71 and ends at amino acid 337. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 3.10E-17. The domain starts at amino acid 887 and ends at amino acid 935. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, has a PH domain, (PFAM profile accession # PF00169), identified with P_score 1.70E-16. The domain starts at amino acid 956 and ends at amino acid 1074. The profile has a length of 85 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 85.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, has a CNH domain, (PFAM profile accession # PF00780), identified with P_score 1.50E-12. The domain starts at amino acid 1100 and ends at amino acid 1380. The profile has a length of 378 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 378.

DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, has a Protein kinase C terminal domain, (PFAM profile accession # PF00433), identified with P_score 2.00E-08. The domain starts at amino acid 351 and ends at amino acid 366. The profile has a length of 70 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 16 to "profile end" residue number 31.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 5.50E-74. The domain starts at amino acid 389 and ends at amino acid 535. The profile has a length of 294 amino

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 149.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 5.50E-74. The domain starts at amino acid 560 and ends at amino acid 662. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 158 to "profile end" residue number 294.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69, has a PDZ domain, (PFAM profile accession # PF00595), identified with P_score 3.70E-09. The domain starts at amino acid 972 and ends at amino acid 1054. The profile has a length of 84 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 79.

MAST205, SEQ ID NO: 4, SEQ ID NO: 70, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 7.90E-80. The domain starts at amino acid 512 and ends at amino acid 785. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MAST205, SEQ ID NO: 4, SEQ ID NO: 70, has a PDZ domain (Also known as DHR or GLGF)., (PFAM profile accession # PF00595), identified with P_score 2.20E-10. The domain starts at amino acid 1104 and ends at amino acid 1191. The profile has a length of 83 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 83.

MASTL, SEQ ID NO: 5, SEQ ID NO: 71, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.20E-73. The domain starts at amino acid 35 and ends at amino acid 310. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MASTL, SEQ ID NO: 5, SEQ ID NO: 71, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.20E-73. The domain starts at amino acid 739 and ends at amino acid 834. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 149 to "profile end" residue number 278.

MASTL, SEQ ID NO: 5, SEQ ID NO: 71, has a Protein kinase C terminal domain, (PFAM profile accession # PF00433), identified with P_score 4.60E-07. The domain starts at amino acid 835 and ends at amino acid 863. The profile has a length of 70 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 31.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.60E-82. The domain starts at amino acid 355 and ends at amino acid 614. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 4.40E-46. The domain starts at amino acid 172 and ends at amino acid 222. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 4.40E-46. The domain starts at amino acid 246 and ends at amino acid 295. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, has a Protein kinase C terminal domain, (PFAM profile accession # PF00433), identified with P_score 1.80E-41. The domain

-164-

starts at amino acid 615 and ends at amino acid 681. The profile has a length of 70 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 70.

H19102, SEQ ID NO: 7, SEQ ID NO: 73, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.20E-64. The domain starts at amino acid 146 and ends at amino acid 398. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MSK1, SEQ ID NO: 8, SEQ ID NO: 74, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.60E-182. The domain starts at amino acid 49 and ends at amino acid 318. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MSK1, SEQ ID NO: 8, SEQ ID NO: 74, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.60E-182. The domain starts at amino acid 427 and ends at amino acid 687. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 2 to "profile end" residue number 278.

MSK1, SEQ ID NO: 8, SEQ ID NO: 74, has a Protein kinase C terminal domain, (PFAM profile accession # PF00433), identified with P_score 2.40E-21. The domain starts at amino acid 319 and ends at amino acid 382. The profile has a length of 70 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 70.

YANK3, SEQ ID NO: 9, SEQ ID NO: 75, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.80E-71. The domain starts at amino acid 93 and ends at amino acid 345. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 287.

MARK2, SEQ ID NO: 10, SEQ ID NO: 76, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.30E-100. The domain starts at amino acid 53 and ends at amino acid 304. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

MARK2, SEQ ID NO: 10, SEQ ID NO: 76, has a Kinase associated domain 1, (PFAM profile accession # PF02149), identified with P_score 3.00E-21. The domain starts at amino acid 738 and ends at amino acid 787. The profile has a length of 50 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 50.

MARK2, SEQ ID NO: 10, SEQ ID NO: 76, has a UBA/TS-N domain, (PFAM profile accession # PF00627), identified with P_score 0.000003. The domain starts at amino acid 324 and ends at amino acid 363. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

NuaK2, SEQ ID NO: 11, SEQ ID NO: 77, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 8.00E-94. The domain starts at amino acid 97 and ends at amino acid 347. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

BRSK2, SEQ ID NO: 12, SEQ ID NO: 78, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.20E-97. The domain starts at amino acid 19 and ends at amino acid 270. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MARK4, SEQ ID NO: 13, SEQ ID NO: 79, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 7.70E-104. The domain starts at amino acid 59 and ends at amino acid 310. The profile has a length of 278 amino

-166-

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MARK4, SEQ ID NO: 13, SEQ ID NO: 79, has a Kinase associated domain 1, (PFAM profile accession # PF02149), identified with P_score 1.30E-15. The domain starts at amino acid 703 and ends at amino acid 752. The profile has a length of 50 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 50.

MARK4, SEQ ID NO: 13, SEQ ID NO: 79, has a UBA domain, (PFAM profile accession # PF00627), identified with P_score 6.30E-11. The domain starts at amino acid 330 and ends at amino acid 368. The profile has a length of 41 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 41.

DCAMKL2, SEQ ID NO: 14, SEQ ID NO: 80, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.70E-97. The domain starts at amino acid 394 and ends at amino acid 651. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

PIM2, SEQ ID NO: 15, SEQ ID NO: 81, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.40E-71. The domain starts at amino acid 132 and ends at amino acid 386. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

PIM3, SEQ ID NO: 16, SEQ ID NO: 82, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 9.90E-80. The domain starts at amino acid 40 and ends at amino acid 293. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.10E-78. The domain starts at amino acid 25 and ends at amino acid 293. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CKIL2, SEQ ID NO: 18, SEQ ID NO: 84, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 8.50E-33. The domain starts at amino acid 21 and ends at amino acid 276. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 265.

PCTAIRE3, SEQ ID NO: 19, SEQ ID NO: 85, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.20E-87. The domain starts at amino acid 50 and ends at amino acid 331. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

PFTAIRE2, SEQ ID NO: 20, SEQ ID NO: 86, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 4.40E-80. The domain starts at amino acid 103 and ends at amino acid 387. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

ERK7, SEQ ID NO: 21, SEQ ID NO: 87, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 4.80E-90. The domain starts at amino acid 13 and ends at amino acid 323. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CKIIa-rs, SEQ ID NO: 22, SEQ ID NO: 88, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.20E-89. The domain starts at amino acid 39 and ends at amino acid 324. The profile has a length of 278 amino

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

DYRK4, SEQ ID NO: 23, SEQ ID NO: 89, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 4.00E-64. The domain starts at amino acid 506 and ends at amino acid 802. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

HIPK1, SEQ ID NO: 24, SEQ ID NO: 90, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 6.20E-58. The domain starts at amino acid 190 and ends at amino acid 518. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

HIPK4, SEQ ID NO: 25, SEQ ID NO: 91, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.10E-58. The domain starts at amino acid 11 and ends at amino acid 347. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

BIKE, SEQ ID NO: 26, SEQ ID NO: 92, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.50E-38. The domain starts at amino acid 51 and ends at amino acid 314. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 8.80E-70. The domain starts at amino acid 519 and ends at amino acid 783. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, has a Armadillo/beta-catenin-like repeat, (PFAM profile accession # PF00514), identified with P_score 0.009707. The domain starts at amino acid 198 and ends at amino acid 238. The profile has a length of 40 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 40.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, has a Armadillo/beta-catenin-like repeat, (PFAM profile accession # PF00514), identified with P_score 0.009707. The domain starts at amino acid 239 and ends at amino acid 279. The profile has a length of 40 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 40.

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, has a Armadillo/beta-catenin-like repeat, (PFAM profile accession # PF00514), identified with P_score 0.009707. The domain starts at amino acid 280 and ends at amino acid 320. The profile has a length of 40 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 40.

pNEK5, SEQ ID NO: 28, SEQ ID NO: 94, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 9.10E-87. The domain starts at amino acid 61 and ends at amino acid 316. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

NEK1, SEQ ID NO: 29, SEQ ID NO: 95, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.50E-89. The domain starts at amino acid 4 and ends at amino acid 258. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

NEK3, SEQ ID NO: 30, SEQ ID NO: 96, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 5.60E-92. The domain starts at amino acid 4 and ends at amino acid 257. The profile has a length of 278 amino

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

SGK069, SEQ ID NO: 31, SEQ ID NO: 97, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.80E-40. The domain starts at amino acid 62 and ends at amino acid 325. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 263.

SGK110, SEQ ID NO: 32, SEQ ID NO: 98, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.70E-39. The domain starts at amino acid 98 and ends at amino acid 359. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 273.

NRBP2, SEQ ID NO: 33, SEQ ID NO: 99, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.00E-24. The domain starts at amino acid 38 and ends at amino acid 313. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CNK, SEQ ID NO: 34, SEQ ID NO: 100, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.60E-91. The domain starts at amino acid 62 and ends at amino acid 314. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CNK, SEQ ID NO: 34, SEQ ID NO: 100, has a POLO box duplicated region., (PFAM profile accession # PF00659), identified with P_score 9.70E-35. The domain starts at amino acid 470 and ends at amino acid 533. The profile has a length of 77 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 77.

CNK, SEQ ID NO: 34, SEQ ID NO: 100, has a POLO box duplicated region., (PFAM profile accession # PF00659), identified with P_score 9.70E-35. The domain starts at amino acid 567 and ends at amino acid 637. The profile has a length of 77 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 77.

SCYL2, SEQ ID NO: 35, SEQ ID NO: 101, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 8.00E-13. The domain starts at amino acid 32 and ends at amino acid 327. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 7.40E-42. The domain starts at amino acid 81 and ends at amino acid 686. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

TLK1, SEQ ID NO: 37, SEQ ID NO: 103, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 4.70E-71. The domain starts at amino acid 477 and ends at amino acid 755. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

SGK071, SEQ ID NO: 38, SEQ ID NO: 104, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 7.60E-26. The domain starts at amino acid 28 and ends at amino acid 296. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 27 to "profile end" residue number 278.

SK516, SEQ ID NO: 39, SEQ ID NO: 105, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.50E-44. The domain starts at amino acid 652 and ends at amino acid 915. The profile has a length of 278 amino

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

H85389, SEQ ID NO: 40, SEQ ID NO: 106, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.90E-60. The domain starts at amino acid 69 and ends at amino acid 397. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

Wee1b, SEQ ID NO: 41, SEQ ID NO: 107, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.10E-49. The domain starts at amino acid 212 and ends at amino acid 486. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 272.

Wnk2, SEQ ID NO: 42, SEQ ID NO: 108, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 6.60E-63. The domain starts at amino acid 181 and ends at amino acid 439. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.00E-85. The domain starts at amino acid 1242 and ends at amino acid 1507. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.10E-88. The domain starts at amino acid 468 and ends at amino acid 731. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

Pak4 (Mus musculus), SEQ ID NO: 45 SEQ ID NO: 111, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 5.00E-86. The domain starts at amino acid 323 and ends at amino acid 574. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

Pak4, SEQ ID NO: 45 SEQ ID NO: 111, has a P21-Rho-binding domain, (PFAM profile accession # PF00786), identified with P_score 3.20E-12. The domain starts at amino acid 11 and ends at amino acid 69. The profile has a length of 64 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 64.

STLK6-rs, SEQ ID NO: 46 SEQ ID NO: 112, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 2.60E-33. The domain starts at amino acid 58 and ends at amino acid 369. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 14 to "profile end" residue number 278.

MAP2K2, SEQ ID NO: 47 SEQ ID NO: 113, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.20E-58. The domain starts at amino acid 72 and ends at amino acid 369. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 6.70E-63. The domain starts at amino acid 796 and ends at amino acid 1061. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 272.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 46 and ends at amino acid 103. The profile has a length of 45 amino

-174-

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 143 and ends at amino acid 202. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 239 and ends at amino acid 303. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 336 and ends at amino acid 393. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 426 and ends at amino acid 483. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 517 and ends at amino acid 572. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

CCK4, SEQ ID NO: 48 SEQ ID NO: 114, has a Immunoglobulin domain, (PFAM profile accession # PF00047), identified with P_score 1.00E-61. The domain starts at amino acid 606 and ends at amino acid 666. The profile has a length of 45 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 45.

LMR1, SEQ ID NO: 49 SEQ ID NO: 115, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.10E-46. The domain starts at amino acid 125 and ends at amino acid 395. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 294.

RYK, SEQ ID NO: 50 SEQ ID NO: 116, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 3.10E-81. The domain starts at amino acid 330 and ends at amino acid 596. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 276.

RYK, SEQ ID NO: 50 SEQ ID NO: 116, has a WIF domain, (PFAM profile accession # PF02019), identified with P_score 3.30E-91. The domain starts at amino acid 66 and ends at amino acid 194. The profile has a length of 132 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 132.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.00E-41. The domain starts at amino acid 1886 and ends at amino acid 2138. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 8 to "profile end" residue number 272.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 983 and ends at amino acid 1004. The profile has a length of 23 amino

-176-

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1012 and ends at amino acid 1035. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23. '

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1036 and ends at amino acid 1058. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1084 and ends at amino acid 1103. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1108 and ends at amino acid 1129. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1130 and ends at amino acid 1153. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1174 and ends at amino acid 1196. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1197 and ends at amino acid 1218. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1221 and ends at amino acid 1244. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1246 and ends at amino acid 1268. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, has a Leucine Rich Repeat, (PFAM profile accession # PF00560), identified with P_score 2.10E-34. The domain starts at amino acid 1269 and ends at amino acid 1293. The profile has a length of 23 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 23.

pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.70E-87. The domain starts at amino acid 124 and ends at amino acid 398. The profile has a length of 294 amino

acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 292.

pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, has a SH3 domain, (PFAM profile accession # PF00018), identified with P_score 2.00E-14. The domain starts at amino acid 45 and ends at amino acid 100. The profile has a length of 58 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 5 to "profile end" residue number 58.

KSR, SEQ ID NO: 53 SEQ ID NO: 119, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.40E-31. The domain starts at amino acid 591 and ends at amino acid 731. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 147.

KSR, SEQ ID NO: 53 SEQ ID NO: 119, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.40E-31. The domain starts at amino acid 753 and ends at amino acid 792. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 163 to "profile end" residue number 195.

KSR, SEQ ID NO: 53 SEQ ID NO: 119, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 0.008623. The domain starts at amino acid 348 and ends at amino acid 391. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

KSR, SEQ ID NO: 53 SEQ ID NO: 119, has a MYND finger, (PFAM profile accession # PF01753), identified with P_score 1.311685. The domain starts at amino acid 360 and ends at amino acid 377. The profile has a length of 43 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 21.

-179-

KSR2, SEQ ID NO: 54 SEQ ID NO: 120, has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 6.90E-40. The domain starts at amino acid 698 and ends at amino acid 957. The profile has a length of 294 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 289.

KSR2, SEQ ID NO: 54 SEQ ID NO: 120, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 0.000127. The domain starts at amino acid 445 and ends at amino acid 488. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, has a Diacylglycerol kinase catalytic domain, (PFAM profile accession # PF00781), identified with P_score 2.50E-09. The domain starts at amino acid 132 and ends at amino acid 278. The profile has a length of 159 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 159.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a Diacylglycerol kinase accessory domain, (PFAM profile accession # PF00609), identified with P_score 3.30E-129. The domain starts at amino acid 582 and ends at amino acid 762. The profile has a length of 190 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 190.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a Diacylglycerol kinase catalytic domain, (PFAM profile accession # PF00781), identified with P_score 1.20E-71. The domain starts at amino acid 438 and ends at amino acid 562. The profile has a length of 159 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 159.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 5.00E-28. The domain starts at amino acid 245 and ends at amino acid 294. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a Phorbol esters/diacylglycerol binding domain (C1 domain), (PFAM profile accession # PF00130), identified with P_score 5.00E-28. The domain starts at amino acid 310 and ends at amino acid 358. The profile has a length of 51 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 51.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a EF hand, (PFAM profile accession # PF00036), identified with P_score 4.10E-17. The domain starts at amino acid 153 and ends at amino acid 181. The profile has a length of 29 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 29.

DGK-beta, SEQ ID NO: 56 SEQ ID NO: 122, has a EF hand, (PFAM profile accession # PF00036), identified with P_score 4.10E-17. The domain starts at amino acid 198 and ends at amino acid 226. The profile has a length of 29 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 29.

IP6K1, SEQ ID NO: 57 SEQ ID NO: 123, did not have a recognizable protein domain.

YAB1, SEQ ID NO: 58 SEQ ID NO: 124, has a ABC1 family, (PFAM profile accession # PF03109), identified with P_score 1.20E-42. The domain starts at amino acid 318 and ends at amino acid 434. The profile has a length of 124 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 124.

BRD2, SEQ ID NO: 62 SEQ ID NO: 128, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 4.90E-91. The domain starts at amino acid 79 and ends at amino acid 168. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRD2, SEQ ID NO: 62 SEQ ID NO: 128, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 4.90E-91. The domain starts at amino acid 352 and ends at amino acid 441. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRD3, SEQ ID NO: 63, SEQ ID NO: 129, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 6.50E-87. The domain starts at amino acid 39 and ends at amino acid 128. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRD3, SEQ ID NO: 63, SEQ ID NO: 129, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 6.50E-87. The domain starts at amino acid 315 and ends at amino acid 403. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRD4, SEQ ID NO: 64, SEQ ID NO: 130, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 1.80E-90. The domain starts at amino acid 63 and ends at amino acid 152. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRD4, SEQ ID NO: 64, SEQ ID NO: 130, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 1.80E-90. The domain starts at amino acid 356 and ends at amino acid 445. The profile has a length of 92 amino acids. The

regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRDT, SEQ ID NO: 65, SEQ ID NO: 131, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 7.50E-86. The domain starts at amino acid 32 and ends at amino acid 121. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

BRDT, SEQ ID NO: 65, SEQ ID NO: 131, has a Bromodomain, (PFAM profile accession # PF00439), identified with P_score 7.50E-86. The domain starts at amino acid 275 and ends at amino acid 364. The profile has a length of 92 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 92.

ZC1, SEQ ID NO: 66, SEQ ID NO: 132 has a Protein kinase domain, (PFAM profile accession # PF00069), identified with P_score 1.4E-91. The domain starts at amino acid 25 and ends at amino acid 289. The profile has a length of 278 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 278.

ZC1, SEQ ID NO: 66, SEQ ID NO: 132 also has a CNH domain, (PFAM profile accession # PF00780), identified with P_score 9.2E-131. The domain starts at amino acid 1066 and ends at amino acid 1372. The profile has a length of 378 amino acids. The regions of the profile that recognized the domain within the protein were from "profile start" residue number 1 to "profile end" residue number 378.

IV. BIOLOGICAL SIGNIFICANCE, APPLICATIONS, AND CLINICAL RELEVANCE

For each protein kinase in this application, we provide a classification of the protein class and family to which it belongs, a summary of non-catalytic protein motifs, and a chromosomal location. This information can be used to suggest potential function, regulation or therapeutic utility for each of the proteins. Amplification of chromosomal region can be associated with various cancers. For amplicons discussed

-183-

in this application, the source of information was Knuutila, et al (Knuutila S, Björkqvist A-M, Autio K, Tarkkanen M, Wolf M, Monni O, Szymanska J, Larramendy ML, Tapper J, Pere H, El-Rifai W, Hemmer S, Wasenius V-M, Vidgren V & Zhu Y: DNA copy number amplifications in human neoplasms. Review of comparative genomic hybridization studies. Am J Pathol 152: 1107-1123, 1998. http: //www.helsinki.fi/lgl_www/CMG.html).

The kinase classification and protein domains often reflect pathways, cellular roles, or mechanisms of up- or down-stream regulation. Also disease-relevant genes often occur in families of related genes. For example if one member of a kinase family functions as an oncogene, a tumor suppressor, or has been found to be disrupted in an immune, neurologic, cardiovascular, or metabolic disorder, frequently other family members may play a related role.

I.   BIOLOGICAL AND POTENTIAL CLINICAL IMPLICATIONS OF THE NOVEL PROTEIN KINASES

AGC Group

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, MAST3, SEQ ID NO: 3, SEQ ID NO: 69, MAST205, SEQ ID NO: 4, SEQ ID NO: 70, MASTL, SEQ ID NO: 5, SEQ ID NO: 71, PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, H19102, SEQ ID NO: 7, SEQ ID NO: 73, MSK1, SEQ ID NO: 8, SEQ ID NO: 74, YANK3, SEQ ID NO: 9, SEQ ID NO: 75 are members of the AGC group of protein kinases. The AGC group of protein kinases includes as its major prototypes protein kinase C (PKC), cAMP-dependent protein kinases (PKA), the G protein-coupled receptor kinases [(ARK and rhodopsin kinase (GRK1)] as well as p70S6K and AKT.

The human CRIK protein and nucleic acid are described in this patent. By PCR of a mouse primary keratinocyte cDNA library, Di Cunto et al. (1998) identified murine CRIK (citron Rho-interacting kinase), belonging to the myotonic dystrophy kinase (see 605377) family. Murine CRIK can be expressed as at least 2 isoforms, one of which encompasses the previously reported form of citron in almost its entirety. The long form of murine CRIK is a 240-kD protein in which the kinase domain is

followed by the sequence of citron. The short murine form, CRIK-SK (short kinase), is an approximately 54-kD protein that consists mostly of the kinase domain. CRIK and CRIK-SK proteins are capable of phosphorylating exogenous substrates as well as of autophosphorylation, when tested by in vitro kinase assays after expression into COS-7 cells. Murine CRIK kinase activity is increased several-fold by coexpression of constitutively active Rho, while active Rac has more limited effects. Kinase activity of the endogenous CRIK is indicated by in vitro kinase assays after immunoprecipitation with antibodies recognizing the citron moiety of the protein. When expressed in keratinocytes, full-length CRIK, but not CRIK-SK, localizes into corpuscular cytoplasmic structures and elicits recruitment of actin into these structures. The CRIK protein contains a kinase domain, a coiled-coil domain, a leucine-rich domain, a Rho-Rac binding domain, a zinc finger region, a pleckstrin homology domain, and a putative SH3-binding domain. Di Cunto, F.; Calautti, E.; Hsiao, J.; Ong, L.; Topley, G.; Turco, E.; Dotto, G. P. : Citron Rho-interacting kinase, a novel tissue-specific ser/thr kinase encompassing the Rho-Rac-binding protein citron. *J. Biol. Chem.* 273: 29706-29711, 1998.

The human DMPK2 protein and nucleic acid are described in this patent. The homolog DMPK1 is associated with myotonic dystrophy (DM), is a multisystem disorder and the most common form of muscular dystrophy in adults. One form of the disorder (Dystrophia Myotonica 1, DM1; 160900) is caused by an expanded CTG repeat in the 3-prime untranslated region of the dystrophia myotonica protein kinase gene (DMPK1; 605377) on 19q13. A CTG repeat in DMPK1 is transcribed and is located in the 3-prime untranslated region of an mRNA that is expressed in tissues affected by myotonic dystrophy. The polypeptide encoded by this mRNA is a member of the protein kinase family. Since the triplet repeat sequence is within a gene that has a sequence similar to protein kinases, Fu et al. (1992) suggested that the gene be referred to as myotonin-protein kinase. Jansen et al. (1992) demonstrated that the brain and heart transcripts of the DM-kinase gene are subject to alternative RNA splicing in both human and mouse. Given the homology between DMPK1 and DMPK2, DMPK2 may be involved in diseases similar to myotonic dystrophy. Fu, Y et al.. *Science* 255: 1256-1258, 1992.

Jansen, G.; et al. **Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs.** *Nature Genet.* 1: 261-266, 1992.

CRIK, SEQ ID NO: 1, SEQ ID NO: 67, and DMPK2, SEQ ID NO: 2, SEQ ID NO: 68 are a members of the DMPK family. These proteins, Dystrophia myotonica-protein kinases, may play a role in muscle contraction; trinucleotide repeat expansion mutations in the 3' untranslated region of DMPK are associated with myotonic dystrophy. These genes may be involved in diseases of the muscle or nerves.

MAST3, SEQ ID NO: 3, SEQ ID NO: 69, MAST205, SEQ ID NO: 4, SEQ ID NO: 70, and MASTL, SEQ ID NO: 5, SEQ ID NO: 71, are a members of the MAST family. Mast protein kinases have strong similarity to microtubule associated testis specific serine/threonine protein kinase (mouse Mtssk), which may act in spermatid maturation and microtubule organization. These kinases may be involved in microtubule-associated disease processes, such as tumor cell invasion.

PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, is a member of the PKC family. Protein kinase C (PKC) is a family of enzymes that are physiologically activated by 1,2-diacylglycerol (DAG) and other lipids. To date, 11 different isozymes, alpha, betaI, betaII, gamma, delta, epsilon, nu, lambda(iota), mu, theta and zeta, have been identified. On the basis of their structure and activators, they can be divided into three groups, two of which are activated by DAG or its surrogate, phorbol 12-myristate 13-acetate (PMA). PKC isozymes are remarkably different in number and prevalence in different cell lines and tissues. When activated, the isozymes bind to membrane phospholipids or to receptors that are located in and anchor the enzymes in a subcellular compartment. Some PKCs may also be activated in their soluble form. These enzymes phosphorylate serine and threonine residues on protein substrates, perhaps the best known of which are the myristoylated, alanine-rich C kinase substrate and nuclear lamins A, B and C. The enzymes clearly play a role in signal transduction, and, because of the importance of PMA as a tumor promoter, they are thought to affect some aspect of cell cycling. (See "The sevenfold way of PKC regulation," Liu WS, Heckman CA, Cell Signal, 1998 Sept10(8): 529-42).

H19102, SEQ ID NO: 7, SEQ ID NO: 73, MSK1, SEQ ID NO: 8, SEQ ID NO: 74, are members of the family of S6 kinases with a potential role in cancer, inflammation, as well as other disease conditions. Ribosomal protein S6 protein kinases play important pleotropic functions, among them is a key role in the regulation of mRNA translation during protein biosynthesis (*Eur J Biochem* 2000 Nov; 267(21): 6321-30, Exp Cell Res. 1999 Nov 25;253 (1): 100-9, *Mol Cell Endocrinol* 1999 May 25;151(1-2): 65-77). The phosphorylation of the S6 ribosomal protein by p70S6 has also been implicated in the regulation of cell motility (*Immunol Cell Biol* 2000 Aug;78(4): 447-51) and cell growth (*Prog Nucleic Acid Res Mol Biol* 2000;65: 101-27), and hence, may be important in tumor metastasis, the immune response and tissue repair.

YANK3, SEQ ID NO: 9, SEQ ID NO: 75, is a member of the Protein Kinase superfamily. It is further classified into the AGC group, and the YANK family.

CAMK Group

MARK2, SEQ ID NO: 10, SEQ ID NO: 76, NuaK2, SEQ ID NO: 11, SEQ ID NO: 77, BRSK2, SEQ ID NO: 12, SEQ ID NO: 78, MARK4, SEQ ID NO: 13, SEQ ID NO: 79, DCAMKL2, SEQ ID NO: 14, SEQ ID NO: 80, PIM2, SEQ ID NO: 15, SEQ ID NO: 81, PIM3, SEQ ID NO: 16, SEQ ID NO: 82, and TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, are classified into the CAMK group. The CAMK group of protein kinases includes as its major prototypes the calmodulin-dependent protein kinases, elongation factor-2 kinases, phosphorylase kinase and the Snf1 and cAMP-dependent family of protein kinases.

CK1 Group

CKIL2, SEQ ID NO: 18, SEQ ID NO: 84, is a member of the Protein Kinase superfamily, the CKI group, and the CKIL family. The casein kinase (CK) group of protein kinases includes as its major prototype casein kinaseI (CKI) and case in kinaseII (CKII). Both CKI and CKII are ubiquitous, constitutively-active, second-messenger-independent kinases These highly conserved enzymes exist in multiple isoforms. CKI functions in vesicular trafficking, DNA repair, cell cycle progression

-187-

and cytokinesis (Cell Signal 1998 Nov;10(10): 699-711). CKII functions in cell cycle progression in non-neural cells. CKII has also been implicated in multiple signaling pathways in normal and disease states of the mammalian nervous systems (Prog Neurobiol 2000 Feb;60(3): 211-46).

Other Group

CKIIa-rs, SEQ ID NO: 22, SEQ ID NO: 88, is a member of the Protein Kinase superfamily, the Other group, and the CKII family.

CMGC Group

PCTAIRE3, SEQ ID NO: 19, SEQ ID NO: 85 and PFTAIRE2, SEQ ID NO: 20, SEQ ID NO: 86 belong in the CMGC group, and the CDK family. The CMGC group of protein kinases includes as its major prototypes the cyclin-dependent protein kinases as well as the MAPK kinases family member. The CDK family to which these kinases belong regulates the cell cycle, as well as transcription and other basic cellular processes.

ERK7, SEQ ID NO: 21, SEQ ID NO: 87, is a member of the Protein Kinase superfamily. It is further classified into the CMGC group, and the MAPK family. Member of the MAP kinase family of proteins, which are involved in signal transduction; may interact with MEK family of kinases.

DYRK4, SEQ ID NO: 23, SEQ ID NO: 89, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the DYRK family.

HIPK1, SEQ ID NO: 24, SEQ ID NO: 90, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the DYRK family.

HIPK4, SEQ ID NO: 25, SEQ ID NO: 91, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the DYRK family.

SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, is a member of the Protein Kinase superfamily. It is further classified into the GMGC group, and the SRPK family. Its role is in mRNA splicing.

Other Family

BIKE, SEQ ID NO: 26, SEQ ID NO: 92, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NAK family. Bike (BMP-2-Inducible Kinase) kinase activity impairs osteoblast differentiation in vitro (Kearns AE, et al.,. J Biol Chem 2001 Nov 9;276(45): 42213-8. Since differentiation of osteoblasts is an important step in the progression of bone diseases such as osteoporosis and cancer associated bone degradation, inhibition of Bike may be an excellent means of treating these diseases, as well as others associated with aberrant bone biology.

NEK Family

NEK10, SEQ ID NO: 27, SEQ ID NO: 93, NEK5, SEQ ID NO: 28, SEQ ID NO: 94, NEK1, SEQ ID NO: 29, SEQ ID NO: 95, NEK3, SEQ ID NO: 30, SEQ ID NO: 96, are members of the Protein Kinase superfamily, the Other group, and the NEK family. The prototype for this family, NIMA (never in mitosis, gene A), was originally identified in Aspergillus nidulans as a serine/threonine kinase critical for cell cycle progression. NIMA is specifically required to initiate the cytological aspects of mitosis. Temperature-sensitive mutants of NIMA or overexpression of dominant negative forms of NIMA cause cells to arrest in G2 with uncondensed DNA and interphase microtubules (Osmani, (1991) Cell 67, 283-291). In addition, overexpression of NIMA in fungus as well as in mammalian cells results in the early onset of mitotic events, including chromatin condensation and depolymerization of microtubules (Lu, K. P., and Hunter, T. (1995) *Prog. Cell Cycle Res.* 1, 187-205). The ability of NIMA to functionally regulate mitosis in higher organisms has suggested the existence of a conserved NIMA-like pathway in eukaryotes. However, only in the filamentous ascomycete, Neurospora crassa, and the fission yeast Schizosaccharomyces pombe have functional homologs been identified. Several mammalian Neks have been identified. These typically contain 40-50% sequence identity, which is confined to the catalytic domain. Positional cloning studies revealed Nek1 as the gene that is altered in polycystic kidney disease, although its precise function remains unknown (Upadhya, P.,. (2000) *Proc. Natl. Acad. Sci.*

U. S. A. **97**, 217-221). Nek2 represents the best characterized mammalian Nek. Nek2 displays cell-cycle dependent expression similar to NIMA, both being most abundant at the onset of mitosis (Fry, A. M., (1995) *J. Biol. Chem.* **270**, 12899-12905). Endogenous Nek2 associates with centrosomes, and overexpression of active Nek2 in cells causes a pronounced splitting of centrosomes, required for G2/M transition. Nek2 phosphorylates a centrosomal coiled-coil protein, c-Nap1, and also associates with protein phosphatase 1 (Helps, N. R., (2000) *Biochem. J.* **349**, 509-518). These findings suggest that Nek2 contributes to proper centrosomal function. Characterization of Nek9 has recently been published (Holland, PM et al, J. Biol. Chem., Vol. 277, Issue 18, 16229-16240, May 3, 2002). The novel NEK genes described in this application may play roles in cell-cycle regulation, protein synthesis, changes in cell morphology and regulation of protein sorting.

These genes are classified within the NKF1 family: SGK069, SEQ ID NO: 31, SEQ ID NO: 97, and SGK110, SEQ ID NO: 32, SEQ ID NO: 98, are members of the Protein Kinase superfamily, classified into the Other group, and the NKF1 family.

NRBP2, SEQ ID NO: 33, SEQ ID NO: 99, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the NRBP family. This family is releated to the WNK family of kinases, and like the WNK family, may be involved in hypertension.

CNK, SEQ ID NO: 34, SEQ ID NO: 100, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the PLK family. CNK seems to be required in a step between RAS and RAF or in parallel to RAF, and its function is required for normal cell proliferation and differentiation (PNAS, Therrien, M., et al, Vol. 96, Issue 23, 13259-13263, November 9, 1999). Its role in Ras signalling may implicate it in aberrant signaling associated with cancer, inflammation or CNS disorders.

SCYL2, SEQ ID NO: 35, SEQ ID NO: 101, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the SCY1 family.

TLK1, SEQ ID NO: 37, SEQ ID NO: 103, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the TLK family.

SGK071, SEQ ID NO: 38, SEQ ID NO: 104, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Unique family.

SK516, SEQ ID NO: 39, SEQ ID NO: 105, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Unique family.

H85389, SEQ ID NO: 40, SEQ ID NO: 106, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the ULK family. It is related to hedgehog signaling.

Wee1b, SEQ ID NO: 41, SEQ ID NO: 107, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the WEE family.

Wnk2, SEQ ID NO: 42, SEQ ID NO: 108, is a member of the Protein Kinase superfamily. It is further classified into the Other group, and the Wnk family. Wnk2 belongs to the same family as Wnk1 and Wnk4, which have been shown to be involved in human hypertension (Wilson FH, et al Science, 2001 Aug 10;293(5532): 1030). Wnk1 and Wnk4 cause pseudohypoaldosteronism type II, a Mendelian trait featuring hypertension, increased renal salt reabsorption, and impaired K+ and H+ excretion. Disease-causing mutations in WNK1 are large intronic deletions that increase WNK1 expression. The mutations in WNK4 are missense, which cluster in a short, highly conserved segment of the encoded protein. Both proteins localize to the distal nephron, a kidney segment involved in salt, K+, and pH homeostasis. WNK1 is cytoplasmic, whereas WNK4 localizes to tight junctions. The WNK kinases and their associated signaling pathway(s) may offer new targets for the development of antihypertensive drugs. Based on its similarity to Wnk1 and Wnk4, Wnk2 may play a role in human hypertension.

STE Group

The STE group of protein kinases represent key regulators of multiple signal transduction pathways important in cell proliferation, survival, differentiation and

response to cellular stress. The STE group of protein kinases includes as its major prototypes the NEK kinases as well as the STE11 and STE20 family of sterile protein kinases. MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, is a member of the STE11 family; Pak5_m, SEQ ID NO: 45 SEQ ID NO: 111, is a member of the STE20 family; STLK6-rs, SEQ ID NO: 46 SEQ ID NO: 112, is a member of the STE20 family; MAP2K2, SEQ ID NO: 47 SEQ ID NO: 113, is a member of the STE7 family. Based on the similarity to STE family members, these novel kinases may participate in cell cycle regulation.

Tyrosine Kinase Group

The tyrosine kinase group encompass both cytoplasmic (e.g. src) as well as transmembrane receptor tyrosine kinases (e.g. EGF receptor). These kinases play a pivotal role in the signal transduction processes that mediate cell proliferation, differentiation and apoptosis. Three genes are classified as tyrosine kinases: CCK4, SEQ ID NO: 48 SEQ ID NO: 114, is classified into the TK group, and the CCK4 family; LMR1, SEQ ID NO: 49 SEQ ID NO: 115, classified into the TK group, and the Lmr family; and RYK, SEQ ID NO: 50 SEQ ID NO: 116, is classified into the TK group, and the Ryk family.

Tyrosine Kinase-Like (TKL) Group

The TKL family represents protein kinases that are more closely related to tyrosine kinases than to serine-threonine kinases. The TKL family consists of the IRAK, LISK, LRRK, MLK, RAF/KSR and STKR sub-families (Manning, G, et al, The Human Kinome, submitted to Science, June 2002; see also www.kinase.com for kinase classification). LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, is classified into the TKL group, and the LRRK family; MLK4, SEQ ID NO: 52 SEQ ID NO: 118, is classified into the TKL group, and the MLK family; KSR, SEQ ID NO: 53 SEQ ID NO: 119, is classified into the TKL group, and the RAF family; KSR2, SEQ ID NO: 54 SEQ ID NO: 120, is classified into the TKL group, and the RAF family.

Lipid Kinase Superfamily

KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, and DGK-beta, SEQ ID NO: 56
SEQ ID NO: 122, are members of the Lipid Kinase superfamily and the DAG/DGK
family. Diacylglycerol kinases (DGKs) phosphorylate the second-messenger
diacylglycerol (DAG) to phosphatidic acid (PA). The family of DGKs is well
conserved among most species. Nine mammalian isotypes have been identified, and
are classified into five subgroups based on their primary structure. DGKs contain a
conserved catalytic domain and an array of other conserved motifs that are likely to
play a role in lipid-protein and protein-protein interactions in various signalling
pathways dependent on DAG and/or PA production. DGK is therefore believed to be
activated at the (plasma) membrane where DAG is generated. Some isotypes are
found associated with and/or regulated by small GTPases of the Rho family. Others
are (also) found in the nucleus, in association with other regulatory enzymes of the
phosphoinositide cycle, and have an effect on cell cycle progression. Most DGK
isotypes show high expression in the brain, often in distinct brain regions, suggesting
that each individual isotype has a unique function. (see "Properties and functions of
diacylglycerol kinases," van Blitterswijk WJ; Cell Signal 2000 Oct;12(9-10): 595-
605).

IP6K1, SEQ ID NO: 57 SEQ ID NO: 123, is a member of the Lipid Kinase
superfamily. It is further classified into the Inositol kinase group, and the IP6K
family (J. Biol. Chem., Vol. 276, Issue 44, 40998-41004, November 2, 2001).
Signaling through the inositol phosphate pathway involves a series of kinases and
phosphatases that phosphorylate and dephosphorylate the large number of soluble
inositol polyphosphates known to exist in eukaryotic cells (Shears, S. B. (1991)
Pharmacol. Ther. 49, 79-104). A branch point in this pathway occurs with the
production of inositol 1,3,4-trisphosphate (Ins(1,3,4)P3)1, resulting from the
hydrolysis of inositol 1,3,4,5-tetrakisphosphate (Ins(1,3,4,5)P4) by one of the
numerous inositol polyphosphate 5-phosphatase isozymes. Ins(1,3,4)P3 can be
dephosphorylated by specific phosphatases, resulting ultimately in the generation of
myo-inositol, or it can be phosphorylated further, resulting in the formation of higher
phosphorylated forms of inositol. Inositol 1,3,4-trisphosphate 5/6-kinase (5/6-kinase)

-193-

phosphorylates Ins(1,3,4)P3 to form both inositol 1,3,4,6-tetrakisphosphate
(Ins(1,3,4,6)P4) and Ins(1,3,4,5)P4. Ins(1,3,4,6)P4 is the first intermediate in the
pathway leading to the formation of the higher phosphorylated inositols including
other inositol tetrakisphosphate isomers, inositol 1,3,4,5,6-pentakisphosphate (InsP5),
inositol hexakisphosphate (InsP6), and the pyrophosphate forms of inositol (Safrany,
S. T., et al. (1999) *Biol. Chem.* **380**, 945-951). IP6K1, SEQ ID NO: 57 SEQ ID NO:
123 may play a role in signalling pathways mediated by phosphoinositol molecules,
such as cancer, inflammation and CNS diseases.

Atypical Group

ABC1 family

YAB1, SEQ ID NO: 58 SEQ ID NO: 124, AF052122, SEQ ID NO: 59 SEQ ID NO:
125, and AAF23326, SEQ ID NO: 60 SEQ ID NO: 126 are members of the ABC1
family. ABC1 is an anciently-conserved family of atypical kinases. The family has
four members in human, five in Drosophila, and three each in C. elegans and S.
cerevisiae. There is weak sequence and structural similarity between ABC1 family
members and eukaryotic protein kinases (see Novel Families of Putative Protein
Kinases in Bacteria and Archaea: Evolution of the Eukaryotic Protein Kinase
Superfamily, CJ Leonared, et al, Genome Research, 8: 1038-1047, 1998). Some
family members are localized to the nucleus or the mitochondrion, and may function
as novel chaperonins and in energy metabolism. Human family members may serve
as targets for disrupting metabolism of cancer cells, for conditions where folding and
turnover of proteins is misregulated, or where disruption of protein folding or
turnover may have a therapeutic effect, as has been seen recently with the use of
proteasome inhibitors to treat a range of cancers.

Rio family

SGK493, SEQ ID NO: 61 SEQ ID NO: 127, is a member of the atypical PK
superfamily, and the RIO1 family. Rio is an anciently-conserved family of atypical
kinases. Three Rio genes are present in the human genome, with distinct orthologs in
fly and worm, and homologs in fungi, archeal bacteria and plants. Rio kinases have
weak protein and structural similarity to eukaryotic protein kinases, and biochemical

kinase activity has recently been shown for the Rio1 family member in S. cerevisiae (Angermayr et al, (2002) Molecular Microbiology 44(2): 309-24). Rio1 is required for proper cell cycle and cell division, and for mRNA processing. Both family members in yeast (Rio1 and Rio2) are essential genes null mutants are lethal. Emericella nidulans sudD is another member of the family and is also involved in cell cycle and chromosome segregation. These conserved functions indicate that human members of this family may play critical roles in cell cycle and constitute tractable targets for cancer therapies.

BRD family

BRD2, SEQ ID NO: 62 SEQ ID NO: 128, BRD3, SEQ ID NO: 63, SEQ ID NO: 129, BRD4, SEQ ID NO: 64, SEQ ID NO: 130, and BRDT, SEQ ID NO: 65, SEQ ID NO: 131, are members of the atypical protein kinase superfamily, belonging to the BRD sub-family This family consists of 4 human members, with a single ortholog in Drosophila and in C. elegans. This phylogenetic footprint indicates that the family plays an essential role in metazoan animals, and has been expanded to serve more specialized or expanded functions in humans. All family members contain two bromodomains, thought to be involved in chromosome biology, and an additional conserved region which bears weak sequence and structural similarity to the eukaryotic protein kinase domain. The Drosophila ortholog, fsh is involved in homeotic gene function and chromosomal imprinting. One of the human family members, BRD2/RING3 has been shown to have protein kinase activity. (Denis GV, et al, RING3 kinase transactivates promoters of cell cycle regulatory genes through E2F.Cell Growth Differ. 2000 Aug;11(8): 417-24). BRD2 expression is elevated in certain human leukemias, is localized to the nucleus and is required for induction of expression of a number of cell cycle genes. This data, and the bromodomains found in other family members indicate that all family members may be involved in control of cell cycle, chromosome function and oncogenic transformation.

EXAMPLE 3: Isolation of cDNAs Encoding Mammalian Protein Kinases

Materials and Methods

Identification of novel clones

Total RNAs are isolated using the Guanidine Salts/Phenol extraction protocol of
Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, *Anal. Biochem.* 162, 156
(1987)) from primary human tumors, normal and tumor cell lines, normal human
tissues, and sorted human hematopoietic cells. These RNAs are used to generate
single-stranded cDNA using the Superscript Preamplification System (GIBCO BRL,
Gaithersburg, MD; Gerard, GF *et al.* (1989), *FOCUS* 11, 66) under conditions
recommended by the manufacturer. A typical reaction uses 10 μg total RNA with 1.5
μg oligo(dT)$_{12-18}$ in a reaction volume of 60 μL. The product is treated with RNaseH
and diluted to 100 μL with H_2O. For subsequent PCR amplification, 1-4 μL of this
sscDNA is used in each reaction.

Degenerate oligonucleotides are synthesized on an Applied Biosystems 3948 DNA
synthesizer using established phosphoramidite chemistry, precipitated with ethanol
and used unpurified for PCR. These primers are derived from the sense and antisense
strands of conserved motifs within the catalytic domain of several protein kinases.
Degenerate nucleotide residue designations are: N = A, C, G, or T; R = A or G; Y = C
or T; H = A, C or T not G; D = A, G or T not C; S = C or G; and W = A or T.

PCR reactions are performed using degenerate primers applied to multiple single-
stranded cDNAs. The primers are added at a final concentration of 5 μM each to a
mixture containing 10 mM TrisHCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM
each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA
Polymerase (Perkin-Elmer/Cetus), and 1-4 μL cDNA. Following 3 min denaturation
at 95 °C, the cycling conditions are 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1
min 45 s for 35 cycles. PCR fragments migrating between 300-350 bp are isolated
from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the
pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

-196-

Colonies are selected for mini plasmid DNA-preparations using Qiagen columns and the plasmid DNA is sequenced using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. *et al.*, *J.Mol.Biol.* 215: 403-10).

Additional PCR strategies are employed to connect various PCR fragments or ESTs using exact or near exact oligonucleotide primers. PCR conditions are as described above except the annealing temperatures are calculated for each oligo pair using the formula: Tm = 4(G+C)+2(A+T).

Isolation of cDNA clones

Human cDNA libraries are probed with PCR or EST fragments corresponding to kinase-related genes. Probes are ^{32}P-labeled by random priming and used at $2x10^6$ cpm/mL following standard techniques for library screening. Pre-hybridization (3 h) and hybridization (overnight) are conducted at 42 oC in 5X SSC, 5X Denhart's solution, 2.5% dextran sulfate, 50 mM Na_2PO_4/$NaHPO_4$, pH 7.0, 50% formamide with 100 mg/mL denatured salmon sperm DNA. Stringent washes are performed at 65 °C in 0.1X SSC and 0.1% SDS. DNA sequencing was carried out on both strands using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer.

EXAMPLE 4: Expression Analysis of Mammalian Protein Kinases

Materials and Methods

Northern blot analysis

Northern blots are prepared by running 10 µg total RNA isolated from 60 human
tumor cell lines (such as HOP-92, EKVX, NCI-H23, NCI-H226, NCI-H322M, NCI-
H460, NCI-H522, A549, HOP-62, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8,
IGROV1, SK-OV-3, SNB-19, SNB-75, U251, SF-268, SF-295, SF-539, CCRF-CEM,
K-562, MOLT-4, HL-60, RPMI 8226, SR, DU-145, PC-3, HT-29, HCC-2998, HCT-
116, SW620, Colo 205, HTC15, KM-12, UO-31, SN12C, A498, CaKi1, RXF-393,
ACHN, 786-0, TK-10, LOX IMVI, Malme-3M, SK-MEL-2, SK-MEL-5, SK-MEL-
28, UACC-62, UACC-257, M14, MCF-7, MCF-7/ADR RES, Hs578T, MDA-MB-
231, MDA-MB-435, MDA-N, BT-549, T47D), from human adult tissues (such as
thymus, lung, duodenum, colon, testis, brain, cerebellum, cortex, salivary gland, liver,
pancreas, kidney, spleen, stomach, uterus, prostate, skeletal muscle, placenta,
mammary gland, bladder, lymph node, adipose tissue), and 2 human fetal normal
tissues (fetal liver, fetal brain), on a denaturing formaldehyde 1.2% agarose gel and
transferring to nylon membranes.

Filters are hybridized with random primed [α^{32}P]dCTP-labeled probes synthesized
from the inserts of several of the kinase genes. Hybridization is performed at 42 °C
overnight in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured
herring sperm DNA with 1-2 x 10^6 cpm/mL of ^{32}P-labeled DNA probes. The filters
are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed on a Molecular Dynamics
phosphorimager.

Quantitative PCR analysis

RNA is isolated from a variety of normal human tissues and cell lines. Single
stranded cDNA is synthesized from 10 µg of each RNA as described above using the
Superscript Preamplification System (GibcoBRL). These single strand templates are
then used in a 25 cycle PCR reaction with primers specific to each clone. Reaction
products are electrophoresed on 2% agarose gels, stained with ethidium bromide and

photographed on a UV light box. The relative intensity of the STK-specific bands were estimated for each sample.

DNA Array Based Expression Analysis

Plasmid DNA array blots are prepared by loading 0.5 µg denatured plasmid for each kinase on a nylon membrane. The $[\gamma^{32}P]dCTP$ labeled single stranded DNA probes are synthesized from the total RNA isolated from several human immune tissue sources or tumor cells (such as thymus, dendrocytes, mast cells, monocytes, B cells (primary, Jurkat, RPMI8226, SR), T cells (CD8/CD4+, TH1, TH2, CEM, MOLT4), K562 (megakaryocytes). Hybridization is performed at 42 °C for 16 hours in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured herring sperm DNA with 10^6 cpm/mL of $[\gamma^{32}P]dCTP$ labeled single stranded probe. The filters are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed for quantitative analysis on a Molecular Dynamics phosphorimager.

EXAMPLE 5: Protein Kinase Gene Expression

Vector Construction

Materials and Methods

Expression Vector Construction

Expression constructs are generated for some of the human cDNAs including: a) full-length clones in a pCDNA expression vector; b) a GST-fusion construct containing the catalytic domain of the novel kinase fused to the C-terminal end of a GST expression cassette; and c) a full-length clone containing a Lys to Ala (K to A) mutation at the predicted ATP binding site within the kinase domain, inserted in the pCDNA vector.

The "K to A" mutants of the kinase might function as dominant negative constructs, and will be used to elucidate the function of these novel STKs.

EXAMPLE 6: Generation of Specific Immunoreagents to Protein Kinases

Materials and Methods

Specific immunoreagents are raised in rabbits against KLH- or MAP-conjugated synthetic peptides corresponding to isolated kinase polypeptides. C-terminal peptides were conjugated to KLH with glutaraldehyde, leaving a free C-terminus. Internal peptides were MAP-conjugated with a blocked N-terminus. Additional immunoreagents can also be generated by immunizing rabbits with the bacterially expressed GST-fusion proteins containing the cytoplasmic domains of each novel PTK or STK.

The various immune sera are first tested for reactivity and selectivity to recombinant protein, prior to testing for endogenous sources.

Western blots

Proteins in SDS PAGE are transferred to immobilon membrane. The washing buffer is PBST (standard phosphate-buffered saline pH 7.4 + 0.1% Triton X-100). Blocking and antibody incubation buffer is PBST +5% milk. Antibody dilutions varied from 1: 1000 to 1: 2000.

EXAMPLE 7: Recombinant Expression and Biological Assays for Protein Kinases

Materials and Methods

Transient Expression of Kinases in Mammalian Cells

The pcDNA expression plasmids (10 μg DNA/100 mm plate) containing the kinase constructs are introduced into 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells are harvested in 0.5 mL solubilization buffer (20 mM HEPES, pH 7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM $MgCl_2$, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 μg/mL aprotinin). Sample aliquots are resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 6% acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific

binding is blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using the various anti-peptide or anti-GST-fusion specific antisera.

In Vitro Kinase Assays

Three days after transfection with the kinase expression constructs, a 10 cm plate of 293 cells is washed with PBS and solubilized on ice with 2 mL PBSTDS containing phosphatase inhibitors (10 mM NaHPO$_4$, pH 7.25, 150 mM NaCl, 1% Triton X-100, 0.5% deoxycholate, 0.1% SDS, 0.2% sodium azide, 1 mM NaF, 1 mM EGTA, 4 mM sodium orthovanadate, 1% aprotinin, 5 μg/mL leupeptin). Cell debris was removed by centrifugation (12000 x g, 15 min, 4 °C) and the lysate was precleared by two successive incubations with 50 μL of a 1: 1 slurry of protein A sepharose for 1 hour each. One-half mL of the cleared supernatant was reacted with 10 μL of protein A purified kinase-specific antisera (generated from the GST fusion protein or antipeptide antisera) plus 50 μL of a 1: 1 slurry of protein A-sepharose for 2 hr at 4 °C. The beads were then washed 2 times in PBSTDS, and 2 times in HNTG (20 mM HEPES, pH 7.5/150 mM NaCl, 0,1% Triton X-100, 10% glycerol).

The immunopurified kinases on sepharose beads are resuspended in 20 μL HNTG plus 30 mM MgCl$_2$, 10 mM MnCl$_2$, and 20 μCi [α^{32}P]ATP (3000 Ci/mmol). The kinase reactions are run for 30 min at room temperature, and stopped by addition of HNTG supplemented with 50 mM EDTA. The samples are washed 6 times in HNTG, boiled 5 min in SDS sample buffer and analyzed by 6% SDS-PAGE followed by autoradiography. Phosphoamino acid analysis is performed by standard 2D methods on ^{32}P-labeled bands excised from the SDS-PAGE gel.

Similar assays are performed on bacterially expressed GST-fusion constructs of the kinases.

EXAMPLE 8a: Chromosomal Localization of Protein Kinases
(Table 5)

Materials and Methods

Chromosomal location can identify candidate targets for a tumor amplicon or a tumor-suppressor locus. Summaries of prevalent tumor amplicons are available in the literature, and can identify tumor types to experimentally be confirmed to contain amplified copies of a kinase gene which localizes to an adjacent region. Several sources were used to find information about the chromosomal localization of each of the genes described in this patent. Materials and Methods

Several sources were used to find information about the chromosomal localization of each of the genes described in this patent. First, the Celera Browser was used to map the genes. A second source was through BLAT searching of the Human Genome using the University of California, Santa Cruz web tools (http: //genome.ucsc.edu/). Alternatively, the accession number of a genomic contig (identified by BLAST against NRNA) was used to query the Entrez Genome Browser (http: //www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewerHelp.html), and the cytogenetic localization was read from the NCBI data. References for association of the mapped sites with chromosomal amplifications found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998, 152: 1107-1123. Information on mapped positions was also obtained by searching published literature (at NCBI, http: //www.ncbi.nlm.nih.gov/entrez/query.fcgi) for documented association of the mapped position with human disease.

1.   Results

The chromosomal regions for mapped genes are listed Table 5, and are discussed in the section Nucleic Acids above. The chromosomal positions were cross-checked with the Online Mendelian Inheritance in Man database (OMIM, http: //www.ncbi.nlm.nih.gov/htbin-post/Omim)., which tracks genetic information for many human diseases, including cancer. References for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in:

-202-

Knuutila, et al., Am J Pathol, 1998, 152: 1107-1123. A third source of information on mapped positions was searching published literature (at NCBI, http: //www.ncbi.nlm.nih.gov/entrez/query.fcgi) for documented association of the mapped position with human disease.

Several sources were used to find information about the chromosomal localization of each of the genes described in this patent. First, cytogenetic map locations of these contigs were found in the title or text of their Genbank record, or by inspection through the NCBI human genome map viewer (http: //www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch?).

Alternatively, the accession number of a genomic contig (identified by BLAST against NRNA) was used to query the Entrez Genome Browser (http: //www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewerHelp.html), and the cytogenetic localization was read from the NCBI data. A thorough search of available literature for the cytogenetic region is also made using Medline (http: //www.ncbi.nlm.nih.gov/PubMed/medline.html). References for association of the mapped sites with chromosomal amplifications found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998, 152: 1107-1123.

Alternatively, the accession number for the nucleic acid sequence is used to query the Unigene database. The site containing the Unigene search engine is: http: //www.ncbi.nlm.nih.gov/UniGene/Hs.Home.html. Information on map position within the Unigene database is imported from several sources, including the Online Mendelian Inheritance in Man (OMIM, http: //www.ncbi.nlm.nih.gov/Omim/searchomim.html), The Genome Database (http: //gdb.infobiogen.fr/gdb/simpleSearch.html), and the Whitehead Institute human physical map (http: //carbon.wi.mit.edu: 8000/cgi-bin/contig/sts_info?database=release).

Once a cytogenetic region has been identified by one of these approaches, disease association can be established by searching OMIM with the cytogenetic location. OMIM maintains a searchable catalog of cytogenetic map locations organized by

disease. A thorough search of available literature for the cytogenetic region is also made using Medline (http: //www.ncbi.nlm.nih.gov/PubMed/medline.html). As noted above, feferences for association of the mapped sites with chromosomal abnormalities found in human cancer can be found in: Knuutila, et al., An. J Pathol, 1998, 152: 1107-1123.

EXAMPLE 8b: Candidate Single Nucleotide Polymorphisms (SNPs) (Table 3)

Materials and Methods

The most common variations in human DNA are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection. Candidate SNPs for the genes in this patent were identified by blastn searching the nucleic acid sequences against the public database of sequences containing documented SNPs (dbSNP: sequence files were downloaded from ftp: //ncbi.nlm.nih.gov/SNP/human/rs-fasta/ and ftp: //ncbi.nlm.nih.gov/SNP/human/ss-fasta/ and used to create a blast database). dbSNP accession numbers for the SNP-containing sequences are given. SNPs were also identified by comparing several databases of expressed genes (dbEST, NRNA) and genomic sequence (i.e., NRNA) for single basepair mismatches. The results are shown in Table 3. These are candidate SNPs – their actual frequency in the human population was not determined. The code below is standard for representing DNA sequence:

G   = Guanosine

A   = Adenosine

T   = Thymidine

C   = Cytidine

R   = G or A, puRine

Y   = C or T, pYrimidine

K   = G or T, Keto

W   = A or T, Weak (2 H-bonds)

-204-

S   = C or G, Strong (3 H-bonds)

M   = A or C, aMino

B   = C, G or T (i.e., not A)

D   = A, G or T (i.e., not C)

H   = A, C or T (i.e., not G)

V   = A, C or G (i.e., not T)

N   = A, C, G or T, aNy

X   = A, C, G or T

complementary   G A T C R Y W S K M B V D H N X

DNA   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

strands   C T A G Y R S W M K V B H D N X

For example, if two versions of a gene exist, one with a "C" at a given position, and a second one with a "T: at the same position, then that position is represented as a Y, which means C or T.

Results

A single nucleotide polymorphism in CRIK, SEQ ID NO: 1, SEQ ID NO: 67, occurs at nucleotide position 2924. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 958. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss1337340_allelePos=258.

A single nucleotide polymorphism in CRIK, SEQ ID NO: 1, SEQ ID NO: 67, occurs at nucleotide position 3377. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 1109. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1631893_allelePos=310.

A single nucleotide polymorphism in CRIK, SEQ ID NO: 1, SEQ ID NO: 67, occurs at nucleotide position 4085. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 1345. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1631886_allelePos=605.

A single nucleotide polymorphism in DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, occurs at nucleotide position 5050. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1752530_allelePos=201.

A single nucleotide polymorphism in DMPK2, SEQ ID NO: 2, SEQ ID NO: 68, occurs at nucleotide position 1139. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 358. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "G." The dbSNP accession number for this SNP is gnl|dbSNP|ss1754079_allelePos=201.

A single nucleotide polymorphism in MAST3, SEQ ID NO: 3, SEQ ID NO: 69, occurs at nucleotide position 2900. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 955. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1846926_allelePos=432.

A single nucleotide polymorphism in MAST3, SEQ ID NO: 3, SEQ ID NO: 69, occurs at nucleotide position 623. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 196. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "H." The dbSNP accession number for this SNP is gnl|dbSNP|ss88979_allelePos=67.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 2739. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 913. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1363030_allelePos=144.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 25. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 9. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): R / stop. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss133576_allelePos=22.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 5303. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 1768. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): S / F. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1529170_allelePos=51.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 4652. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 1551. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): D/G. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1529101_allelePos=5.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 3590. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 1197. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): K/R. The amino acid at this position in the patent sequence is "K." The dbSNP accession number for this SNP is gnl|dbSNP|ss1529096_allelePos=51.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 156. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 52. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1608593_allelePos=756.

A single nucleotide polymorphism in MAST205, SEQ ID NO: 4, SEQ ID NO: 70, occurs at nucleotide position 162. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 54. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss497486_allelePos=201.

A single nucleotide polymorphism in MASTL, SEQ ID NO: 5, SEQ ID NO: 71, occurs at nucleotide position 3831. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1363_allelePos=40.

A single nucleotide polymorphism in PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, occurs at nucleotide position 1840. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 558. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "N." The dbSNP accession number for this SNP is gnl|dbSNP|ss1000395_allelePos=101.

A single nucleotide polymorphism in PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, occurs at nucleotide position 1239. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 358. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): T/I. The amino acid at this position in the patent sequence is "I." The dbSNP accession number for this SNP is gnl|dbSNP|ss1472906_allelePos=327.

A single nucleotide polymorphism in PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, occurs at nucleotide position 2288. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1548761_allelePos=51.

A single nucleotide polymorphism in PKC_eta, SEQ ID NO: 6, SEQ ID NO: 72, occurs at nucleotide position 681. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 172. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): H/G. The

amino acid at this position in the patent sequence is "H." The dbSNP accession number for this SNP is gnl|dbSNP|ss1509877_allelePos=51.

A single nucleotide polymorphism in MSK1, SEQ ID NO: 8, SEQ ID NO: 74, occurs at nucleotide position 3186. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2025310_allelePos=201.

A single nucleotide polymorphism in MSK1, SEQ ID NO: 8, SEQ ID NO: 74, occurs at nucleotide position 3658. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1530678_allelePos=5.

A single nucleotide polymorphism in MSK1, SEQ ID NO: 8, SEQ ID NO: 74, occurs at nucleotide position 3769. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1530679_allelePos=51.

A single nucleotide polymorphism in MSK1, SEQ ID NO: 8, SEQ ID NO: 74, occurs at nucleotide position 3432. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1530677_allelePos=51.

A single nucleotide polymorphism in MSK1, SEQ ID NO: 8, SEQ ID NO: 74, occurs at nucleotide position 3779. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1530680_allelePos=51.

A single nucleotide polymorphism in YANK3, SEQ ID NO: 9, SEQ ID NO: 75, occurs at nucleotide position 1852. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss18125_allelePos=101.

A single nucleotide polymorphism in YANK3, SEQ ID NO: 9, SEQ ID NO: 75, occurs at nucleotide position 1895. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1517863_allelePos=5.

A single nucleotide polymorphism in YANK3, SEQ ID NO: 9, SEQ ID NO: 75, occurs at nucleotide position 2021. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1517886_allelePos=51.

A single nucleotide polymorphism in MARK2, SEQ ID NO: 10, SEQ ID NO: 76, occurs at nucleotide position 2570. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 724. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1121403_allelePos=101.

A single nucleotide polymorphism in MARK2, SEQ ID NO: 10, SEQ ID NO: 76, occurs at nucleotide position 2615. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 739. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss1121404_allelePos=101.

A single nucleotide polymorphism in MARK2, SEQ ID NO: 10, SEQ ID NO: 76, occurs at nucleotide position 1641. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 415. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): P/A. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537647_allelePos=51.

A single nucleotide polymorphism in MARK2, SEQ ID NO: 10, SEQ ID NO: 76, occurs at nucleotide position 1547. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 383. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|rs1057176_allelePos=51.

A single nucleotide polymorphism in NuaK2, SEQ ID NO: 11, SEQ ID NO: 77, occurs at nucleotide position 1670. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 538. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1295001_allelePos=93.

A single nucleotide polymorphism in NuaK2, SEQ ID NO: 11, SEQ ID NO: 77, occurs at nucleotide position 1727. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 557. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1295000_allelePos=36.

A single nucleotide polymorphism in MARK4, SEQ ID NO: 13, SEQ ID NO: 79, occurs at nucleotide position 2916. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1967699_allelePos=201.

A single nucleotide polymorphism in MARK4, SEQ ID NO: 13, SEQ ID NO: 79, occurs at nucleotide position 3032. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1967700_allelePos=242.

A single nucleotide polymorphism in MARK4, SEQ ID NO: 13, SEQ ID NO: 79, occurs at nucleotide position 1699. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 561. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1967693_allelePos=201.

A single nucleotide polymorphism in MARK4, SEQ ID NO: 13, SEQ ID NO: 79, occurs at nucleotide position 3092. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1512875_allelePos=51.

A single nucleotide polymorphism in PIM2, SEQ ID NO: 15, SEQ ID NO: 81, occurs at nucleotide position 630. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 210. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "E." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525746_allelePos=5.

-213-

A single nucleotide polymorphism in PIM2, SEQ ID NO: 15, SEQ ID NO: 81, occurs at nucleotide position 1749. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1525747_allelePos=51.

A single nucleotide polymorphism in PIM2, SEQ ID NO: 15, SEQ ID NO: 81, occurs at nucleotide position 1990. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1525754_allelePos=51.

A single nucleotide polymorphism in PIM3, SEQ ID NO: 16, SEQ ID NO: 82, occurs at nucleotide position 2057. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1548948_allelePos=51.

A single nucleotide polymorphism in PIM3, SEQ ID NO: 16, SEQ ID NO: 82, occurs at nucleotide position 1269. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 278. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss1511148_allelePos=51.

A single nucleotide polymorphism in PIM3, SEQ ID NO: 16, SEQ ID NO: 82, occurs at nucleotide position 2362. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1511284_allelePos=51.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 1203. The polymorphism results in the following SNP:

R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 196. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q/R. The amino acid at this position in the patent sequence is "Q." The dbSNP accession number for this SNP is gnl|dbSNP|ss1975997_allelePos=201.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 152. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1588747_allelePos=749.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 141. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1588746_allelePos=738.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 238. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1211997_allelePos=524.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 84. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss934600_allelePos=307.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 281. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the

. following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1747635_allelePos=2506.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 236. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1747634_allelePos=2461.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 136. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2056655_allelePos=355.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 22. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss45790_allelePos=479.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 243. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2061784_allelePos=1157.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 226. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2061783_allelePos=1140.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 47. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1990388_allelePos=1229.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 158. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1911350_allelePos=370.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 77. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1909793_allelePos=506.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 137. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1908525_allelePos=1475.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 44. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1897673_allelePos=1677.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 11. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the

following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1857878_allelePos=1145.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 223. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1816570_allelePos=267.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 85. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1799649_allelePos=306.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 280. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1732367_allelePos=496.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 97. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1729216_allelePos=408.

A single nucleotide polymorphism in TSSK4, SEQ ID NO: 17, SEQ ID NO: 83, occurs at nucleotide position 148. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1684407_allelePos=417.

A single nucleotide polymorphism in CKIL2, SEQ ID NO: 18, SEQ ID NO: 84, occurs at nucleotide position 3889. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 1208. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): H / D. The amino acid at this position in the patent sequence is "H." The dbSNP accession number for this SNP is gnl|dbSNP|ss1551913_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1103. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 318. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537202_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1008. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 287. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): S / R. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537192_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 663. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 172. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): R / stop. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537165_allelePos=51.

-219-

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1428. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1537238_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 194. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 15. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "V." The dbSNP accession number for this SNP is gnl|dbSNP|ss5453_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1200. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 351. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): M / V. The amino acid at this position in the patent sequence is "V." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537218_allelePos=5.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1181. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 344. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537216_allelePos=51.

A single nucleotide polymorphism in CKIIar, SEQ ID NO: 22, SEQ ID NO: 88, occurs at nucleotide position 1104. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the

-220-

following region (UTR or amino acid number): 319. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): M / L. The amino acid at this position in the patent sequence is "M." The dbSNP accession number for this SNP is gnl|dbSNP|ss1537203_allelePos=51.

A single nucleotide polymorphism in DYRK4, SEQ ID NO: 23, SEQ ID NO: 89, occurs at nucleotide position 269. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 90. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): R / H. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss88136_allelePos=155.

A single nucleotide polymorphism in HIPK1, SEQ ID NO: 24, SEQ ID NO: 90, occurs at nucleotide position 4114. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss12250_allelePos=101.

A single nucleotide polymorphism in BIKE, SEQ ID NO: 26, SEQ ID NO: 92, occurs at nucleotide position 1606. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 468. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "Q." The dbSNP accession number for this SNP is gnl|dbSNP|ss1509438_allelePos=51.

A single nucleotide polymorphism in NEK10, SEQ ID NO: 27, SEQ ID NO: 93, occurs at nucleotide position 1149. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 325. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): T/S. The

amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss727804_allelePos=20.

A single nucleotide polymorphism in NEK10, SEQ ID NO: 27, SEQ ID NO: 93, occurs at nucleotide position 1849. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 558. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "G." The dbSNP accession number for this SNP is gnl|dbSNP|ss1891242_allelePos=201.

A single nucleotide polymorphism in NEK10, SEQ ID NO: 27, SEQ ID NO: 93, occurs at nucleotide position 2967. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 931. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N/S. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1325417_allelePos=338.

A single nucleotide polymorphism in NEK1, SEQ ID NO: 29, SEQ ID NO: 95, occurs at nucleotide position 5063. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1520330_allelePos=51.

A single nucleotide polymorphism in NEK1, SEQ ID NO: 29, SEQ ID NO: 95, occurs at nucleotide position 4848. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1520329_allelePos=51.

A single nucleotide polymorphism in NEK3, SEQ ID NO: 30, SEQ ID NO: 96, occurs at nucleotide position 1854. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the

following region (UTR or amino acid number): 3' UTR. The dbSNP accession
number for this SNP is gnl|dbSNP|ss3403_allelePos=2.

A single nucleotide polymorphism in SGK069, SEQ ID NO: 31, SEQ ID NO: 97,
occurs at nucleotide position 1001. The polymorphism results in the following SNP:
S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the
following region (UTR or amino acid number): 298. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): P / A. The
amino acid at this position in the patent sequence is "A." The dbSNP accession
number for this SNP is gnl|dbSNP|ss1317629_allelePos=393.

A single nucleotide polymorphism in SGK069, SEQ ID NO: 31, SEQ ID NO: 97,
occurs at nucleotide position 323. The polymorphism results in the following SNP:
Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the
following region (UTR or amino acid number): 72. The SNP has the following effect
on the coding sequence of the gene (amino acid change or silent): R / C. The amino
acid at this position in the patent sequence is "R." The dbSNP accession number for
this SNP is gnl|dbSNP|ss1688815_allelePos=201.

A single nucleotide polymorphism in SGK110, SEQ ID NO: 32, SEQ ID NO: 98,
occurs at nucleotide position 299. The polymorphism results in the following SNP:
W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the
following region (UTR or amino acid number): 1. The SNP has the following effect
on the coding sequence of the gene (amino acid change or silent): M / L. The amino
acid at this position in the patent sequence is "M." The dbSNP accession number for
this SNP is gnl|dbSNP|ss767141_allelePos=201.

A single nucleotide polymorphism in SGK110, SEQ ID NO: 32, SEQ ID NO: 98,
occurs at nucleotide position 985. The polymorphism results in the following SNP:
R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the
following region (UTR or amino acid number): 229. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss827468_allelePos=20.

A single nucleotide polymorphism in SGK110, SEQ ID NO: 32, SEQ ID NO: 98, occurs at nucleotide position 640. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 114. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss661406_allelePos=201.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2219. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 681. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): L / F. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525084_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2047. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 623. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "F." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525076_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2040. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 621. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q / R. The

amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525074_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2035. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 619. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "Y." The dbSNP accession number for this SNP is gnl|dbSNP|rs1050422_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2021. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 615. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): I / L. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525069_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2014. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 612. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q / H. The amino acid at this position in the patent sequence is "H." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525066_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2029. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 617. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "G." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525072_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2017. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 613. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "F." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525068_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2016. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 613. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Y / F. The amino acid at this position in the patent sequence is "F." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525067_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2001. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 608. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / S. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525064_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 1999. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 607. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "G." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525063_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 1996. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 606. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525062_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 1969. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 597. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525061_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2044. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 622. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "E." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525075_allelePos=51.

A single nucleotide polymorphism in SRPK2, SEQ ID NO: 36, SEQ ID NO: 102, occurs at nucleotide position 2023. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 615. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

-227-

amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525072_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2174. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 646. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): V/D. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515391_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2489. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 751. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N/S. The amino acid at this position in the patent sequence is "N." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515399_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2515. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 760. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515400_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2358. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 707. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "E." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515395_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2294. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 686. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Y / F. The amino acid at this position in the patent sequence is "F." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515394_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2229. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 664. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "V." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515393_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2014. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 593. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515384_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 1137. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 300. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "I." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515380_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 3279. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1515413_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 3142. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1515412_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 2488. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 751. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / Y. The amino acid at this position in the patent sequence is "N." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515398_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 1711. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 492. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): D/Y. The amino acid at this position in the patent sequence is "Y." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515382_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 1730. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the

-230-

following region (UTR or amino acid number): 498. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): S/Y. The
amino acid at this position in the patent sequence is "Y." The dbSNP accession
number for this SNP is gnl|dbSNP|ss1515383_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103,
occurs at nucleotide position 1083. The polymorphism results in the following SNP:
M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the
following region (UTR or amino acid number): 282. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): E/D. The
amino acid at this position in the patent sequence is "E." The dbSNP accession
number for this SNP is gnl|dbSNP|ss1515377_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103,
occurs at nucleotide position 1647. The polymorphism results in the following SNP:
Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the
following region (UTR or amino acid number): 470. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): silent. The
amino acid at this position in the patent sequence is "H." The dbSNP accession
number for this SNP is gnl|dbSNP|ss1515381_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103,
occurs at nucleotide position 1092. The polymorphism results in the following SNP:
R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the
following region (UTR or amino acid number): 285. The SNP has the following
effect on the coding sequence of the gene (amino acid change or silent): silent. The
amino acid at this position in the patent sequence is "K." The dbSNP accession
number for this SNP is gnl|dbSNP|ss1515379_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103,
occurs at nucleotide position 1035. The polymorphism results in the following SNP:
Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the
following region (UTR or amino acid number): 266. The SNP has the following

-231-

effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515376_allelePos=51.

A single nucleotide polymorphism in TLK1, SEQ ID NO: 37, SEQ ID NO: 103, occurs at nucleotide position 951. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 238. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss1515375_allelePos=51.

A single nucleotide polymorphism in Wnk2, SEQ ID NO: 42, SEQ ID NO: 108, occurs at nucleotide position 7079. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2899_allelePos=78.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 2716. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 906. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): I / V. The amino acid at this position in the patent sequence is "I." The dbSNP accession number for this SNP is gnl|dbSNP|ss1317910_allelePos=285.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 6227. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1146242_allelePos=109.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 5560. The polymorphism results in the following SNP:

R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1286358_allelePos=101.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 3187. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 1063. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1146312_allelePos=101.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 6015. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1146243_allelePos=101.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 2416. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 806. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / D. The amino acid at this position in the patent sequence is "N." The dbSNP accession number for this SNP is gnl|dbSNP|ss1146310_allelePos=101.

A single nucleotide polymorphism in MAP3K1, SEQ ID NO: 43, SEQ ID NO: 109, occurs at nucleotide position 1284. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 428. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss1146300_allelePos=101.

-233-

A single nucleotide polymorphism in MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, occurs at nucleotide position 247. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 83. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q/E. The amino acid at this position in the patent sequence is "E." The dbSNP accession number for this SNP is gnl|dbSNP|ss1394913_allelePos=101.

A single nucleotide polymorphism in MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, occurs at nucleotide position 2485. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1617_allelePos=49.

A single nucleotide polymorphism in MAP3K8, SEQ ID NO: 44, SEQ ID NO: 110, occurs at nucleotide position 2298. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1547718_allelePos=51.

A single nucleotide polymorphism in STLK6r, SEQ ID NO: 46 SEQ ID NO: 112, occurs at nucleotide position 487. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 82. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss1483412_allelePos=100.

A single nucleotide polymorphism in Map2K2, SEQ ID NO: 47 SEQ ID NO: 113, occurs at nucleotide position 904. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 219. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

amino acid at this position in the patent sequence is "I." The dbSNP accession number for this SNP is gnl|dbSNP|ss1937135_allelePos=201.

A single nucleotide polymorphism in CCK4, SEQ ID NO: 48 SEQ ID NO: 114, occurs at nucleotide position 3636. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1527472_allelePos=51.

A single nucleotide polymorphism in RYK, SEQ ID NO: 50 SEQ ID NO: 116, occurs at nucleotide position 2875. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss16914_allelePos=101.

A single nucleotide polymorphism in RYK, SEQ ID NO: 50 SEQ ID NO: 116, occurs at nucleotide position 2496. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1525573_allelePos=51.

A single nucleotide polymorphism in RYK, SEQ ID NO: 50 SEQ ID NO: 116, occurs at nucleotide position 851. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 254. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / S. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525514_allelePos=51.

A single nucleotide polymorphism in RYK, SEQ ID NO: 50 SEQ ID NO: 116, occurs at nucleotide position 386. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 99. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / S. The amino

acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1525513_allelePos=51.

A single nucleotide polymorphism in RYK, SEQ ID NO: 50 SEQ ID NO: 116, occurs at nucleotide position 2764. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss16913_allelePos=31.

A single nucleotide polymorphism in LRRK2, SEQ ID NO: 51 SEQ ID NO: 117, occurs at nucleotide position 5425. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 1598. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): E/V. The amino acid at this position in the patent sequence is "V." The dbSNP accession number for this SNP is gnl|dbSNP|ss63276_allelePos=97.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 3597. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057123_allelePos=323.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 3914. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057120_allelePos=201.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 3668. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057122_allelePos=288.

-236-

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 3800. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057121_allelePos=22.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 2580. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 773. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1411720_allelePos=519.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 2611. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 784. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): G/C. The amino acid at this position in the patent sequence is "C." The dbSNP accession number for this SNP is gnl|dbSNP|ss1411719_allelePos=488.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 4193. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057119_allelePos=201.

A single nucleotide polymorphism in pMLK4, SEQ ID NO: 52 SEQ ID NO: 118, occurs at nucleotide position 4309. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2057118_allelePos=201.

A single nucleotide polymorphism in KSR, SEQ ID NO: 53 SEQ ID NO: 119, occurs at nucleotide position 4096. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss100899_allelePos=172.

A single nucleotide polymorphism in KSR2, SEQ ID NO: 54 SEQ ID NO: 120, occurs at nucleotide position 612. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 204. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "T." The dbSNP accession number for this SNP is gnl|dbSNP|ss2005786_allelePos=201.

A single nucleotide polymorphism in KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, occurs at nucleotide position 3769. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2052346_allelePos=499.

A single nucleotide polymorphism in KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, occurs at nucleotide position 3020. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2052345_allelePos=201.

A single nucleotide polymorphism in KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, occurs at nucleotide position 2577. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2052344_allelePos=201.

A single nucleotide polymorphism in KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, occurs at nucleotide position 2391. The polymorphism results in the following SNP:

-238-

R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2052344_allelePos=201.

A single nucleotide polymorphism in KIAA1646, SEQ ID NO: 55 SEQ ID NO: 121, occurs at nucleotide position 4272. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss2052347_allelePos=201.

A single nucleotide polymorphism in IP6K1, SEQ ID NO: 57 SEQ ID NO: 123, occurs at nucleotide position 3669. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1522850_allelePos=51.

A single nucleotide polymorphism in IP6K1, SEQ ID NO: 57 SEQ ID NO: 123, occurs at nucleotide position 2851. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1522846_allelePos=51.

A single nucleotide polymorphism in YAB1, SEQ ID NO: 58 SEQ ID NO: 124, occurs at nucleotide position 2506. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1305707_allelePos=99.

A single nucleotide polymorphism in YAB1, SEQ ID NO: 58 SEQ ID NO: 124, occurs at nucleotide position 1538. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 480. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The

-239-

amino acid at this position in the patent sequence is "F." The dbSNP accession number for this SNP is gnl|dbSNP|ss1529336_allelePos=51.

A single nucleotide polymorphism in SGK493, SEQ ID NO: 61 SEQ ID NO: 127, occurs at nucleotide position 1094. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 349. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): R / G. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1826551_allelePos=201.

A single nucleotide polymorphism in SGK493, SEQ ID NO: 61 SEQ ID NO: 127, occurs at nucleotide position 1690. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 547. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1826528_allelePos=201.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 920. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1425392_allelePos=324.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 1794. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 31. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "K." The dbSNP accession number for this SNP is gnl|dbSNP|ss686785_allelePos=201.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 3510. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 603. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|rs516535_allelePos=201.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 2413. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 238. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): L / F. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss1973307_allelePos=201.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 3199. The polymorphism results in the following SNP: K (G/T). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 500. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): E/stop. The amino acid at this position in the patent sequence is "E." The dbSNP accession number for this SNP is gnl|dbSNP|ss15121_allelePos=101.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 3333. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 544. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "K." The dbSNP accession number for this SNP is gnl|dbSNP|ss13218_allelePos=101.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 4348. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. - The dbSNP accession number for this SNP is gnl|dbSNP|ss12998_allelePos=101.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 3411. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 570. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1550506_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 1344. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1550446_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 4416. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1550446_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 4219. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1523158_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 3342. The polymorphism results in the following SNP:

-242-

R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 547. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "R." The dbSNP accession number for this SNP is gnl|dbSNP|ss1523069_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 811. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 5' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1522874_allelePos=51.

A single nucleotide polymorphism in BRD2, SEQ ID NO: 62 SEQ ID NO: 128, occurs at nucleotide position 2379. The polymorphism results in the following SNP: S (C/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 226. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss18333_allelePos=31.

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 2405. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss575919_allelePos=201.

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 1075. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 312. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "L." The dbSNP accession number for this SNP is gnl|dbSNP|ss630265_allelePos=201.

-243-

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 1975. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 612. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss601346_allelePos=201.

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 1423. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 428. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss634964_allelePos=201.

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 2934. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss17101_allelePos=101.

A single nucleotide polymorphism in BRD3, SEQ ID NO: 63, SEQ ID NO: 129, occurs at nucleotide position 2796. The polymorphism results in the following SNP: Y (C/T). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1527035_allelePos=51.

A single nucleotide polymorphism in BRD4, SEQ ID NO: 64, SEQ ID NO: 130, occurs at nucleotide position 1846. The polymorphism results in the following SNP: R (A/G). The nucleotide in the patent sequence is "G." The SNP occurs within the following region (UTR or amino acid number): 542. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): N / D. The

amino acid at this position in the patent sequence is "D." The dbSNP accession number for this SNP is gnl|dbSNP|ss1512910_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 821. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "A." The SNP occurs within the following region (UTR or amino acid number): 238. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): K / N. The amino acid at this position in the patent sequence is "K." The dbSNP accession number for this SNP is gnl|dbSNP|ss1559581_allelePos=482.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 2976. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 3' UTR. The dbSNP accession number for this SNP is gnl|dbSNP|ss1553268_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 2785. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 893. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q / P. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553264_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 1114. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 336. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): stop / S. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553262_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 1113. The polymorphism results in the following SNP: W (A/T). The nucleotide in the patent sequence is "T." The SNP occurs within the following region (UTR or amino acid number): 336. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Y/S. The amino acid at this position in the patent sequence is "S." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553261_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 2882. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 925. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553267_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 2851. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 915. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): Q/P. The amino acid at this position in the patent sequence is "P." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553266_allelePos=51.

A single nucleotide polymorphism in BRDT, SEQ ID NO: 65, SEQ ID NO: 131, occurs at nucleotide position 2846. The polymorphism results in the following SNP: M (A/C). The nucleotide in the patent sequence is "C." The SNP occurs within the following region (UTR or amino acid number): 913. The SNP has the following effect on the coding sequence of the gene (amino acid change or silent): silent. The amino acid at this position in the patent sequence is "A." The dbSNP accession number for this SNP is gnl|dbSNP|ss1553265_allelePos=51.

EXAMPLE 9: Demonstration Of Gene Amplification By Southern Blotting

Materials and Methods

Nylon membranes are purchased from Boehringer Mannheim. Denaturing solution contains 0.4 M NaOH and 0.6 M NaCl. Neutralization solution contains 0.5 M Tris-HCL, pH 7.5 and 1.5 M NaCl. Hybridization solution contains 50% formamide, 6X SSPE, 2.5X Denhardt's solution, 0.2 mg/mL denatured salmon DNA, 0.1 mg/mL yeast tRNA, and 0.2 % sodium dodecyl sulfate. Restriction enzymes are purchased from Boehringer Mannheim. Radiolabeled probes are prepared using the Prime-it II kit by Stratagene. The beta actin DNA fragment used for a probe template is purchased from Clontech.

Genomic DNA is isolated from a variety of tumor cell lines (such as MCF-7, MDA-MB-231, Calu-6, A549, HCT-15, HT-29, Colo 205, LS-180, DLD-1, HCT-116, PC3, CAPAN-2, MIA-PaCa-2, PANC-1, AsPc-1, BxPC-3, OVCAR-3, SKOV3, SW 626 and PA-1, and from two normal cell lines.

A 10 µg aliquot of each genomic DNA sample is digested with EcoR I restriction enzyme and a separate 10 µg sample is digested with Hind III restriction enzyme. The restriction-digested DNA samples are loaded onto a 0.7% agarose gel and, following electrophoretic separation, the DNA is capillary-transferred to a nylon membrane by standard methods (Sambrook, J. et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory).

EXAMPLE 10: Detection Of Protein-Protein Interaction Through Phage Display

Materials And Methods

Phage display provides a method for isolating molecular interactions based on affinity for a desired bait. cDNA fragments cloned as fusions to phage coat proteins are displayed on the surface of the phage. Phage(s) interacting with a bait are enriched by affinity purification and the insert DNA from individual clones is analyzed.

-247-

T7 Phage Display Libraries

All libraries were constructed in the T7Select1-1b vector (Novagen) according to the manufacturer's directions.

Bait Presentation

Protein domains to be used as baits are generated as C-terminal fusions to GST and expressed in *E. coli*. Peptides are chemically synthesized and biotinylated at the N-terminus using a long chain spacer biotin reagent.

Selection

Aliquots of refreshed libraries (10^{10}-10^{12} pfu) supplemented with PanMix and a cocktail of *E. coli* inhibitors (Sigma P-8465) are incubated for 1-2 hrs at room temperature with the immobilized baits. Unbound phage is extensively washed (at least 4 times) with wash buffer.

After 3-4 rounds of selection, bound phage is eluted in 100 μL of 1% SDS and plated on agarose plates to obtain single plaques.

Identification of insert DNAs

Individual plaques are picked into 25 μL of 10 mM EDTA and the phage is disrupted by heating at 70 °C for 10 min. 2 μL of the disrupted phage are added to 50 μL PCR reaction mix. The insert DNA is amplified by 35 rounds of thermal cycling (94 °C, 50 sec; 50 °C, 1min; 72 °C, 1min).

Composition of Buffer
10x PanMix
5% Triton X-100
10% non-fat dry milk (Carnation)
10 mM EGTA
250 mM NaF
250 μg/mL Heparin (sigma)
250 μg/mL sheared, boiled salmon sperm DNA (sigma)

0.05% Na azide

Prepared in PBS

Wash Buffer

PBS supplemented with:

0.5% NP-40

25 µl g/mL heparin

PCR reaction mix

1.0 mL	10x PCR buffer (Perkin-Elmer, with 15 mM Mg)
0.2 mL	each dNTPs (10 mM stock)
0.1 mL	T7UP primer (15 pmol/µL) GGAGCTGTCGTATTCCAGTC
0.1 mL	T7DN primer (15 pmol/µL)

AACCCCTCAAGACCCGTTTAG

0.2 mL	25 mM MgCl$_2$ or MgSO$_4$ to compensate for EDTA

Q.S. to 10 mL with distilled water

Add 1 unit of Taq polymerase per 50 µL reaction

LIBRARY: T7 Select1-H441

EXAMPLE 26: HUV-EC-C Assay

The following protocol may also be used to measure a compound's activity against PDGF-R, FGF-R, VEGF, aFGF or Flk-1/KDR, all of which are naturally expressed by HUV-EC cells.

DAY 0

1.   Wash and trypsinize HUV-EC-C cells (human umbilical vein endothelial cells, (American Type Culture Collection; catalogue no. 1730 CRL). Wash with Dulbecco's phosphate-buffered saline (D-PBS; obtained from Gibco BRL; catalogue no. 14190-029) 2 times at about 1 ml/10 cm^2 of tissue culture flask. Trypsinize with 0.05% trypsin-EDTA in non-enzymatic cell dissociation solution (Sigma Chemical Company; catalogue no. C-1544). The 0.05% trypsin was made by diluting 0.25% trypsin/1 mM EDTA (Gibco; catalogue no. 25200-049) in the cell dissociation solution. Trypsinize with about 1 ml/25-30 cm^2 of tissue culture flask for

-249-

about 5 minutes at 37 °C. After cells have detached from the flask, add an equal volume of assay medium and transfer to a 50 ml sterile centrifuge tube (Fisher Scientific; catalogue no. 05-539-6).

2.   Wash the cells with about 35 ml assay medium in the 50 ml sterile centrifuge tube by adding the assay medium, centrifuge for 10 minutes at approximately 200 g, aspirate the supernatant, and resuspend with 35 ml D-PBS. Repeat the wash two more times with D-PBS, resuspend the cells in about 1 ml assay medium/15 cm^2 of tissue culture flask. Assay medium consists of F12K medium (Gibco BRL; catalogue no. 21127-014) + 0.5% heat-inactivated fetal bovine serum. Count the cells with a Coulter Counter™ Coulter Electronics, Inc.) and add assay medium to the cells to obtain a concentration of $0.8-1.0 \times 10^5$ cells/ml.

3.   Add cells to 96-well flat-bottom plates at 100 μl/well or $0.8-1.0 \times 10^4$ cells/well; incubate ~24 h at 37 °C, 5% CO_2.

DAY 1

1.   Make up two-fold drug titrations in separate 96-well plates, generally 50 μM on down to 0 μM. Use the same assay medium as mentioned in day 0, step 2, above. Titrations are made by adding 90 μl/well of drug at 200 μM (4X the final well concentration) to the top well of a particular plate column. Since the stock drug concentration is usually 20 mM in DMSO, the 200 μM drug concentration contains 2% DMSO.

Therefore, diluent made up to 2% DMSO in assay medium (F12K + 0.5% fetal bovine serum) is used as diluent for the drug titrations in order to dilute the drug but keep the DMSO concentration constant. Add this diluent to the remaining wells in the column at 60 μl/well. Take 60 μl from the 120 μl of 200 μM drug dilution in the top well of the column and mix with the 60 μl in the second well of the column. Take 60 μl from this well and mix with the 60 μl in the third well of the column, and so on until two-fold titrations are completed. When the next-to-the-last well is mixed, take 60 μl of the 120 μl in this well and discard it. Leave the last well with 60 μl of DMSO/media diluent as a non-drug-containing control. Make 9 columns of titrated drug, enough for triplicate wells each for 1) VEGF (obtained from Pepro Tech Inc.,

-250-

catalogue no. 100-200, 2) endothelial cell growth factor (ECGF) (also known as acidic fibroblast growth factor, or aFGF) (obtained from Boehringer Mannheim Biochemica, catalogue no. 1439 600); or, 3) human PDGF B/B (1276-956, Boehringer Mannheim, Germany) and assay media control. ECGF comes as a preparation with sodium heparin.

2.   Transfer 50 µl/well of the drug dilutions to the 96-well assay plates containing the $0.8-1.0 \times 10^4$ cells/100 µl/well of the HUV-EC-C cells from day 0 and incubate ~2 h at 37 °C, 5% CO_2.

3.   In triplicate, add 50 µl/well of 80 µg/ml VEGF, 20 ng/ml ECGF, or media control to each drug condition. As with the drugs, the growth factor concentrations are 4X the desired final concentration. Use the assay media from day 0, step 2, to make the concentrations of growth factors. Incubate approximately 24 hours at 37 °C, 5% CO_2. Each well will have 50 µl drug dilution, 50 µl growth factor or media, and 100 µl cells, = 200 µl /well total. Thus the 4X concentrations of drugs and growth factors become 1X once everything has been added to the wells.

DAY 2

1.   Add ^{3}H-thymidine (Amersham; catalogue no. TRK-686) at 1 µCi/well (10 µl/well of 100 µCi/ml solution made up in RPMI media + 10% heat-inactivated fetal bovine serum) and incubate ~24 h at 37 °C, 5% CO_2. Note: ^{3}H-thymidine is made up in RPMI media because all of the other applications for which we use the ^{3}H-thymidine involve experiments done in RPMI. The media difference at this step is probably not significant. RPMI was obtained from Gibco BRL, catalogue no. 11875-051.

DAY 3

1.   Freeze plates overnight at -20°C.

DAY 4

1.   Thaw plates and harvest with a 96-well plate harvester (Tomtec Harvester 96[(R)]) onto filter mats (Wallac; catalogue no. 1205-401); read counts on a Wallac Betaplate[(TM)] liquid scintillation counter.

-251-

CONCLUSION

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations that are not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also

-252-

thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

In view of the degeneracy of the genetic code, other combinations of nucleic acids also encode the claimed peptides and proteins of the invention. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acid alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3^{100}, or 5 x 10^{47}, nucleic acid sequences. Thus, a nucleic acid sequence can be modified to form a second nucleic acid sequence, encoding the same polypeptide as encoded by the first nucleic acid sequences, using routine procedures and without undue experimentation. Thus, all possible nucleic acids that encode the claimed peptides and proteins are also fully described herein, as if all were written out in full taking into account the codon usage, especially that preferred in humans. Furthermore, changes in the amino acid sequences of polypeptides, or in the corresponding nucleic acid sequence encoding such polypeptide, may be designed or selected to take place in an area of the sequence where the significant activity of the polypeptide remains unchanged. For example, an amino acid change may take place within a β-turn, away from the active site of the polypeptide. Also changes such as deletions (e.g. removal of a segment of the polypeptide, or in the corresponding nucleic acid sequence encoding such polypeptide, which does not affect the active site) and additions (e.g. addition of more amino acids to the polypeptide sequence without affecting the function of the active site, such as the formation of GST-fusion proteins, or additions in the corresponding nucleic acid sequence encoding such polypeptide without affecting the function of the active site) are also within the scope of the present invention. Such changes to the polypeptides can be performed by those with ordinary skill in the art using routine procedures and without undue experimentation. Thus, all possible nucleic and/or amino acid sequences that can readily be determined not to affect a significant activity of the peptide or protein of the invention are also fully described herein.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

TABLE 1
Description of Open Reading Frames

Gene NAME	Sp	ID#na	ID#aa	Super-family	Group	Family	NA_length	AA_length	ORF Start	ORF End	ORF Length	Physical Status
CRIK	H	1	67	Protein Kinase	AGC	DMPK	8656	2055	51	6218	6168	FL
DMPK2	H	2	68	Protein Kinase	AGC	DMPK	5438	1572	66	4784	4719	Partial
MAST3	H	3	69	Protein Kinase	AGC	MAST	5990	1332	36	4031	3996	Partial
MAST205	H	4	70	Protein Kinase	AGC	MAST	5516	1798	1	5397	5397	Partial
MASTL	H	5	71	Protein Kinase	AGC	MAST	3882	878	967	3603	2637	FL
PKC_eta	H	6	72	Protein Kinase	AGC	PKC	2392	683	407	2458	2052	FL
H19102	H	7	73	Protein Kinase	AGC	RSK	1564	449	188	1537	1350	Partial
MSK1	H	8	74	Protein Kinase	AGC	RSK	3813	802	159	2567	2409	FL
YANK3	H	9	75	Protein Kinase	AGC	YANK	2051	486	70	1530	1461	FL
MARK2	H	10	76	Protein Kinase	CAMK	CAMKL	3063	787	399	2762	2364	Partial
NuaK2	H	11	77	Protein Kinase	CAMK	CAMKL	3463	672	57	2075	2019	FL
BRSK2	H	12	78	Protein Kinase	CAMK	CAMKL	3831	674	25	2049	2025	Partial
MARK4	H	13	79	Protein Kinase	CAMK	CAMKL	3249	752	17	2275	2259	Partial
DCAMKL2	H	14	80	Protein Kinase	CAMK	DCAMKL	2827	766	350	2650	2301	FL
PIM2	H	15	81	Protein Kinase	CAMK	PIM	2186	435	1	1305	1305	Partial
PIM3	H	16	82	Protein Kinase	CAMK	PIM	2405	326	436	1416	981	FL
TSSK4	H	17	83	Protein Kinase	CAMK	TSSK	1710	328	617	1603	987	FL
CKIL2	H	18	84	Protein Kinase	CKI	CKIL	5946	1244	368	4102	3735	FL
PCTAIRE3	H	19	85	Protein Kinase	CMGC	CDK	3229	505	303	1817	1515	FL
PFTAIRE2	H	20	86	Protein Kinase	CMGC	CDK	2250	435	45	1352	1308	FL
ERK7	H	21	87	Protein Kinase	CMGC	MAPK	1906	563	19	1710	1692	FL
CKIalfa	H	22	88	Protein Kinase	Other	CKII	1494	391	150	1325	1176	Partial
DYRK4	H	23	89	Protein Kinase	CMGC	DYRK	2886	921	1	2786	2766	FL
HIPK1	H	24	90	Protein Kinase	CMGC	DYRK	8212	1210	286	3918	3633	FL
HIPK4	H	25	91	Protein Kinase	CMGC	DYRK	3142	616	977	2827	1851	FL
BIKE	H	26	92	Protein Kinase	Other	NAK	3895	1161	203	3688	3486	FL
NEK10	H	27	93	Protein Kinase	Other	NEK	3912	1125	176	3553	3378	FL
pNEK5	H	28	94	Protein Kinase	Other	NEK	2816	889	147	2816	2670	FL
NEK1	H	29	95	Protein Kinase	Other	NEK	5583	1286	493	4353	3861	Partial
NEK3	H	30	96	Protein Kinase	Other	NEK	2326	506	296	1816	1521	Partial
SGK069	H	31	97	Protein Kinase	Other	NKF1	1156	348	110	1156	1047	FL
SGK110	H	32	98	Protein Kinase	Other	NKF1	1853	414	299	1543	1245	FL
NRBP2	H	33	99	Protein Kinase	Other	NRBP	3765	507	282	1805	1524	FL
CNK	H	34	100	Protein Kinase	Other	PLK	2535	646	534	2474	1941	Partial
SCYL2	H	35	101	Protein Kinase	Other	SCY1	5525	933	173	2974	2802	Partial
SRPK2	H	36	102	Protein Kinase	CMGC	SRPK	3715	688	179	2245	2067	FL
TLK1	H	37	103	Protein Kinase	Other	TLK	4321	787	238	2601	2364	Partial
SGK071	H	38	104	Protein Kinase	Other	Unique	2285	632	195	2093	1899	FL
SK51E	H	39	105	Protein Kinase	Other	Unique	7364	829	180	2969	2790	Partial
H85389	H	40	106	Protein Kinase	Other	ULK	1971	401	134	1339	1206	FL
Wee1b	H	41	107	Protein Kinase	Other	WEE	1704	567	1	1704	1704	Partial
Wnk2	H	42	108	Protein Kinase	Other	Wnk	7981	2245	67	6804	6738	Partial
MAP3K1	H	43	109	Protein Kinase	STE	STE11	7026	1511	1	4538	4536	Partial
MAP3K8	H	44	110	Protein Kinase	STE	STE11	2571	735	1	2208	2208	Partial
Pak4_m	M	45	111	Protein Kinase	STE	STE20	1782	593	1	1782	1782	Partial
STLK6-rs	H	46	112	Protein Kinase	STE	STE20	2171	418	242	1498	1257	Partial
MAP2K2	H	47	113	Protein Kinase	STE	STE7	1724	380	248	1390	1143	FL
CCK4	H	48	114	Protein Kinase	TK	CCK4	4232	1070	191	3403	3213	FL
LMR1	H	49	115	Protein Kinase	TK	Lmr	5313	1374	85	4209	4125	FL
RYK	H	50	116	Protein Kinase	TK	Ryk	3663	607	91	1914	1824	Partial
LRRK2	H	51	117	Protein Kinase	TKL	LRRK	9753	2534	633	8237	7605	Partial
pMLK4	H	52	118	Protein Kinase	TKL	MLK	4687	1036	262	3372	3111	FL
KSR	H	53	119	Protein Kinase	TKL	RAF	5913	901	165	2870	2706	Partial
KSR2	H	54	120	Protein Kinase	TKL	RAF	2994	982	1	2949	2949	FL
KIAA1546	H	55	121	Lipid Kinase	DAG kin	DAG kin	4429	537	92	1705	1614	Partial
DGK-beta	H	56	122	Lipid Kinase	DAG kin	DAG kin	4297	804	372	2786	2415	FL
IP6K1	H	57	123	Lipid Kinase	Inositol kinase	IP6K	4461	441	309	1634	1326	Partial
YABI	H	58	124	Atypical PK	Atypical	ABC1	2508	647	99	2042	1944	FL
AF052122	H	59	125	Atypical PK	Atypical	ABC1	5237	591	1	1776	1776	FL
AAF29326	H	60	126	Atypical PK	Atypical	ABC1	1368	455	1	1368	1368	FL
SGK493	H	61	127	Atypical PK	Atypical	RIO1	1832	552	50	1708	1659	FL
BRD2	H	62	128	Atypical PK	BRD	BRD	4693	801	1702	4107	2406	Partial
BRD3	H	63	129	Atypical PK	BRD	BRD	3085	726	140	2320	2181	Partial
BRD4	H	64	130	Atypical PK	BRD	BRD	3149	722	223	2391	2169	Partial
BRDT	H	65	131	Atypical PK	BRD	BRD	3106	947	108	2951	2844	Partial
ZC1	H	66	132	Protein Kinase	STE	STE20	7986	1392	366	4544	4179	FL

255

TABLE 2A
Smith-Waterman Comparison with NCBI Non-redundant Proteins
Page 1 of 2 for Smith-Waterman

Gene NAME	Sp	ID (Gene)	ID (Dlas)	Super-family	Group	Family	AA length	PSCORE	MATCHES	% Identity	% Similar	ACCESSION	DESCRIPTION
CRIK	H	1	67	Protein Kinase	AGC	DMPK	2055	0	1075	98	98	AAG72823	Rho/rac-interacting citron kinase [Mus musculus]
DMPK2	H	2	68	Protein Kinase	AGC	DMPK	1872	2.20E-211	731	45	63	NP_446109	BTK related to the myotonic dystrophy PK [Rattus norvegicus]
MAST3	H	3	69	Protein Kinase	AGC	MAST	1331	0	1287	99	99	BAA25487	[AB011131] KIAA0561 protein [Homo sapiens]
MAST205	H	4	70	Protein Kinase	AGC	MAST	1798	0	1864	99	99	NP_055927	KIAA0807 protein [Homo sapiens]
MASTL	H	5	71	Protein Kinase	AGC	MAST	876	0	876	99	99	NP_116233	Hypothetical protein FLJ14813 [Homo sapiens]
PKC_eta	H	6	72	Protein Kinase	AGC	PKC	683	0	679	99	99	NP_006246	(NM_006255) protein kinase C, eta [Homo sapiens]
H1P102	H	7	73	Protein Kinase	AGC	RSK	449	1.00E-124	269	98	98	BAB71555	Unnamed protein product [Homo sapiens]
MSKL	H	8	74	Protein Kinase	AGC	RSK	802	3.50E-304	767	98	98	NP_004746	Ribosomal protein S6 kinase, polypeptide 5 [Homo sapiens]
YANK3	H	9	75	Protein Kinase	AGC	YANK	455	6.9e-311	444	91	94	AAH08457	Hypothetical serine/threonine protein kinase [Mus musculus]
MARK2	H	10	76	Protein Kinase	CAMK	CAMKL	787	2.50E-299	752	99	99	AAH08771	(BC008771) Similar to ELKL motif kinase [Homo sapiens]
NuaK2	H	11	77	Protein Kinase	CAMK	CAMKL	672	5.10E-280	628	100	100	NP_112214	(NM_030962) hypothetical protein DKFZp434J037 [Homo sapiens]
BRSK2	H	12	78	Protein Kinase	CAMK	CAMKL	674	4.20E-175	602	99	99	CAA07195	Putative serine/threonine protein kinase [Homo sapiens]
MARK4	H	13	79	Protein Kinase	CAMK	CAMKL	752	4.30E-280	751	99	99	AAL33583	MARK4 serine/threonine protein kinase [Homo sapiens]
DCAMKL2	H	14	80	Protein Kinase	CAMK	DCAMKL	766	8.10E-159	513	67	80	O15075	DCAMKL1 (doublecortin-like and CAMK-like 1) [Homo sapiens]
PIM2	H	15	81	Protein Kinase	CAMK	PIM	434	1.40E-145	334	100	100	NP_006969	Pim-2 oncogene; proto-oncogene Pim-2 [Homo sapiens]
PIM3	H	16	82	Protein Kinase	CAMK	PIM	326	9.00E-174	311	85	87	AAH17621	Serine threonine kinase pim3 [Mus musculus]
TSSK4	H	17	83	Protein Kinase	CAMK	TSSK	328	1.60E-69	201	85	94	BAB30483	Putative [Mus musculus]
CKIL2	H	18	84	Protein Kinase	CK0	CKIL	1244	1.60E-298	645	100	100	BAA74870	KIAA0647 protein [Homo sapiens]
PCTAIRE3	H	19	85	Protein Kinase	CMGC	CDK	504	1.50E-220	471	93	93	O07002	Serine/threonine protein kinase PCTAIRE-3 [Homo sapiens]
PFTAIRE2	H	20	86	Protein Kinase	CMGC	CDK	435	8.40E-100	225	88	91	NP_036304	(NM_011074) PFTAIRE protein kinase 1 [Mus musculus]
ERK7	H	21	87	Protein Kinase	CMGC	MAPK	563	1.90E-128	394	87	78	AAD12719.2	Extracellular signal-regulated kinase 2; ERK7 [Rattus norvegicus]
CKIIa-rs	H	22	88	Protein Kinase	Other	CKII	391	8.60E-165	390	99	100	CAA49758	Casein kinase II alpha subunit [Homo sapiens]
DYRK4	H	23	89	Protein Kinase	CMGC	DYRK	821	1.20E-304	628	99	99	Q9NR20	DYRK4 [Homo sapiens]
HIPK1	H	24	90	Protein Kinase	CMGC	DYRK	1210	0	1181	97	99	AAD41592	Myak-L [Mus musculus]
HIPK4	H	25	91	Protein Kinase	CMGC	DYRK	616	0	598	97	98	BAB72060	Hypothetical protein [Macaca fascicularis]
BIKE	H	26	92	Protein Kinase	Other	NAK	1161	7.50E-244	960	82	89	NP_542439	(NM_080709) Bmp2-inducible kinase [Mus musculus]
NEK10	H	27	93	Protein Kinase	Other	NEK	1125	8.80E-165	428	77	90	BAB71395	Unknown (protein for MGC:14652) [Homo sapiens]
pNEK3	H	28	94	Protein Kinase	Other	NEK	889	1.60E-76	180	85	82	P51954	(AK057247) unnamed protein product [Homo sapiens]
NEK1	H	29	95	Protein Kinase	Other	NEK	1286	0	1266	87	82	BAB57794	STK.NEK1 (NimA-related protein kinase 1) [Mus musculus]
NEK3	H	30	96	Protein Kinase	Other	NEK	506	1.80E-202	456	99	99	P51955	KIAA1901 protein [Homo sapiens]
SGK069	H	31	97	Protein Kinase	Other	NKF1	346	7.40E-48	122	42	60	AAK62420	NEK3 (HSPK 36) [Homo sapiens]
SGK110	H	32	98	Protein Kinase	Other	NKF1	414	4.00E-35	110	41	60	871897	p63.7 gastrula-specific [Xenopus laevis]
NRBP2	H	33	99	Protein Kinase	Other	NRBP	507	3.20E-158	300	81	75	NP_037624	Protein Kinase Bsk146 [Danio rerio]
CNK	H	34	100	Protein Kinase	Other	PLK	846	8.60E-238	645	82	90	AAH13699	Nuclear receptor binding protein [Homo sapiens]
SCT2	H	35	101	Protein Kinase	Other	SCY1	893	0	791	89	92	BAA02588	KIAA1380 protein [Homo sapiens]
SRPK2	H	36	102	Protein Kinase	CAGC	SRPK	688	7.80E-183	694	99	99	NP_003129	(NM_003138) SFRS protein kinase 2 [Homo sapiens]
TLK1	H	37	103	Protein Kinase	Other	TLK	787	0	777	88	89	NP_036422	(NM_012290) tousled-like kinase 1 [Homo sapiens]
SGK071	H	38	104	Protein Kinase	Other	Unique	632	0.000001	63	30	50	NP_176653	Hypothetical protein [Arabidopsis thaliana]
SK316	H	39	105	Protein Kinase	Other	Unique	929	5.70E-180	365	100	100	BAA32317	KIAA0472 protein [Homo sapiens]
HBS369	H	40	106	Protein Kinase	Other	ULK	401	2.40E-182	400	98	99	CAC10510.2	Novel protein kinase [Homo sapiens]
Wee1b	H	41	107	Protein Kinase	Other	WEE	567	2.00E-287	541	98	99	AAD04728	Similar to wee1-like protein kinase [Homo sapiens]
Wnk2	H	42	108	Protein Kinase	Other	Wnk	2245	0	1325	98	97	BAB21851	KIAA1760 protein [Homo sapiens]
MAP3K1	H	43	109	Protein Kinase	STE	STE11	1511	0	1459	97	97	Q13233	MEKK 1 [Homo sapiens]
MAP3K8	H	44	110	Protein Kinase	STE	STE11	735	2.80E-92	188	100	100	XP_017343	Hypothetical protein fragment FLJ23074 [Homo sapiens]
Pak4_m	M	45	111	Protein Kinase	STE	STE20	593	2.70E-130	560	92	96	NP_005076	p21-activated kinase 4, effector for Cdc42Hs [Homo sapiens]
STLK6-rs	H	46	112	Protein Kinase	STE	STE20	418	5.90E-222	407	97	98	NP_061041.2	Amyotrophic lateral sclerosis 2 candidate 2 [Homo sapiens]
MAP2K2	H	47	113	Protein Kinase	STE	STE7	381	4.80E-156	353	92	95	NP_109587	p45 (MAP kinase kinase 2) [Homo sapiens]
CCK4	H	48	114	Protein Kinase	TK	CCK4	1070	0	1069	99	99	JC4593	RTK PTK7 precursor [Homo sapiens]
LMR1	H	49	115	Protein Kinase	TK	Lmr	1374	0	1207	100	100	I37690	Apoptosis-associated tyrosine kinase [Homo sapiens]
RYK	H	50	116	Protein Kinase	TK	Lmr	607	3.60E-287	603	99	99	NP_004911	Protein-tyrosine kinase Ryk [Homo sapiens]
LRRK2	H	51	117	Protein Kinase	TKL	LRRK	2554	7.90E-189	463	84	92	NP_060006	RIKEN cDNA 4921613O20 gene [Mus musculus]
pMLK4	H	52	118	Protein Kinase	TKL	MLK	1036	0	1027	89	89	CAC84640	(AJ311788) mixed lineage kinase 4pdta [Homo sapiens]
KSR	H	53	119	Protein Kinase	TKL	RAF	901	3.30E-269	797	88	92	NP_038599	(NM_013571) kinase suppressor of ras [Mus musculus]
KIAA1645	H	54	120	Protein Kinase	TKL	RAF	962	9.60E-119	462	48	62	NP_038599	(NM_013571) kinase suppressor of ras [Mus musculus]
DGK-beta	H	55	121	Lipid Kinase	DAG Kin	DAG Kin	552	0	481	100	100	BAB33316	KIAA1645 protein [Homo sapiens]
YAB1	H	56	122	Lipid Kinase	DAG Kin	DAG Kin	804	0	804	100	100	Q96I77	Diacylglycerol kinase, beta (DGKBETA) [Homo sapiens]
	H	57	123	Atypical PK	Incalloi kinase	IPEK	441	1.00E-257	441	100	100	BAA13393.2	KIAA0263 protein [Homo sapiens]
AFG52122	M	58	124	Atypical PK	Atypical	ABC1	647	3.80E-244	366	100	98	NP_054832	Chaperone, ABC1 activity of bc1 complex like [Homo sapiens]
AAF23326	H	59	125	Atypical PK	Atypical	ABC1	591	1.20E-246	385	100	100	AAH13014	Hypothetical protein [Homo sapiens]
SGK493	H	60	126	Atypical PK	Atypical	RIO1	455	1.40E-304	455	100	100	NP_065154	Hypothetical protein [Homo sapiens]
BRD2	H	61	127	Atypical PK	BRD	BRD	801	2.80E-256	801	100	100	NP_060813	Bromodomain-containing protein 2 [Homo sapiens]
BRD3	H	62	128	Atypical PK	BRD	BRD	726	2.20E-243	801	100	100	NP_005096	Bromodomain-containing protein 3 [Homo sapiens]
BRD4	H	63	129	Atypical PK	BRD	BRD	722	2.60E-212	722	100	100	NP_055114	Bromodomain-containing protein 4 [Homo sapiens]
BRDT	H	64	130	Atypical PK	BRD	BRD	947	0	947	100	100	NP_001717	Testis-specific bromodomain protein [Homo sapiens]
ZG1	H	65	131	Protein Kinase	STE	STE20	1382	0	1203	89	87	NP_032722	NCK interacting kinase; HPK/GCK-like kinase [Mus musculus]

256

TABLE 2B
Smith-Waterman Comparison with NCBI Non-redundant Proteins
Page 2 of 2 for Smith-Waterman

Gene_NAME	Sp	ID#na	ID#aa	Super-family	Group	Family	QUERYSTART	QUERYEND	TARGETSTART	TARGETEND	%QUERY	%TARGET
CRIK	H	1	67	Protein Kinase	AGC	DMPK	1	2055	1	2055	96	96
DMPK2	H	2	68	Protein Kinase	AGC	DMPK	2	1462	4	1588	48	42
MAST3	H	3	69	Protein Kinase	AGC	MAST	39	1331	18	1308	96	98
MAST205	H	4	70	Protein Kinase	AGC	MAST	1	1687	1	1687	93	97
MASTL	H	5	71	Protein Kinase	AGC	MAST	1	878	1	878	99	99
PKC_eta	H	6	72	Protein Kinase	AGC	PKC	1	683	1	682	99	99
H19102	H	7	73	Protein Kinase	AGC	RSK	41	310	1	271	59	98
MSK1	H	8	74	Protein Kinase	AGC	RSK	1	800	1	800	98	97
YANK3	H	9	75	Protein Kinase	AGC	YANK	1	485	1	487	91	90
MARK2	H	10	76	Protein Kinase	CAMK	CAMKL	34	787	1	755	95	99
NuaK2	H	11	77	Protein Kinase	CAMK	CAMKL	45	672	1	628	93	100
BRSK2	H	12	78	Protein Kinase	CAMK	CAMKL	72	674	1	603	89	99
MARK4	H	13	79	Protein Kinase	CAMK	CAMKL	1	752	1	752	99	99
DCAMKL2	H	14	80	Protein Kinase	CAMK	DCAMKL	1	741	1	739	66	69
PIM2	H	15	81	Protein Kinase	CAMK	PIM	101	434	1	334	76	100
PIM3	H	16	82	Protein Kinase	CAMK	PIM	1	326	1	326	95	95
TSSK4	H	17	83	Protein Kinase	CAMK	TSSK	1	328	1	328	85	85
CKIL2	H	18	84	Protein Kinase	CKI	CKIL	600	1244	1	645	51	100
PCTAIRE3	H	19	85	Protein Kinase	CMGC	CDK	1	502	1	472	93	99
PFTAIRE2	H	20	86	Protein Kinase	CMGC	CDK	97	426	129	458	51	47
ERK7	H	21	87	Protein Kinase	CMGC	MAPK	1	560	1	544	68	70
CKIIa-rs	H	22	88	Protein Kinase	Other	CKII	1	391	1	391	99	99
DYRK4	H	23	89	Protein Kinase	CMGC	DYRK	395	921	15	541	57	97
HIPK1	H	24	90	Protein Kinase	CMGC	DYRK	1	1210	1	1210	97	97
HIPK4	H	25	91	Protein Kinase	CMGC	DYRK	1	616	1	616	97	97
BIKE	H	26	92	Protein Kinase	Other	NAK	1	1161	1	1138	82	84
NEK10	H	27	93	Protein Kinase	Other	NEK	698	1125	10	484	38	88
pNEK5	H	28	94	Protein Kinase	Other	NEK	58	333	1	275	20	23
NEK1	H	29	95	Protein Kinase	Other	NEK	1	1286	8	1265	97	99
NEK3	H	30	96	Protein Kinase	Other	NEK	48	506	1	459	90	99
SGK069	H	31	97	Protein Kinase	Other	NKF1	1	348	394	763	99	41
SGK110	H	32	98	Protein Kinase	Other	NKF1	96	359	9	272	26	30
NRBP2	H	33	99	Protein Kinase	Other	NRBP	17	502	44	518	59	56
CNK	H	34	100	Protein Kinase	Other	PLK	1	646	1	646	99	99
SCYL2	H	35	101	Protein Kinase	Other	SCY1	140	933	3	796	84	99
SRPK2	H	36	102	Protein Kinase	CMGC	SRPK	1	688	1	686	99	99
TLK1	H	37	103	Protein Kinase	Other	TLK	1	787	1	787	98	98
SGK071	H	38	104	Protein Kinase	Other	Unique	25	228	1	197	9	10
SK516	H	39	105	Protein Kinase	Other	Unique	565	929	1	365	39	100
H85389	H	40	106	Protein Kinase	Other	ULK	1	401	118	517	99	77
Wee1b	H	41	107	Protein Kinase	Other	WEE	1	559	1	541	95	100
Wnk2	H	42	108	Protein Kinase	Other	Wnk	860	2245	1	1386	61	99
MAP3K1	H	43	109	Protein Kinase	STE	STE11	21	1511	2	1495	96	97
MAP3K8	H	44	110	Protein Kinase	STE	STE11	547	714	1	168	22	100
Pak4_m	M	45	111	Protein Kinase	STE	STE20	1	593	1	591	92	93
STLK6-rs	H	46	112	Protein Kinase	STE	STE20	1	418	1	418	97	97
MAP2K2	H	47	113	Protein Kinase	STE	STE7	2	380	1	380	92	88
CCK4	H	48	114	Protein Kinase	TK	CCK4	1	1070	1	1070	99	99
LMR1	H	49	115	Protein Kinase	TK	Lmr	168	1374	1	1207	87	100
RYK	H	50	116	Protein Kinase	TK	Ryk	1	607	1	607	99	99
LRRK2	H	51	117	Protein Kinase	TKL	LRRK	1990	2534	17	561	18	82
pMLK4	H	52	118	Protein Kinase	TKL	MLK	1	1036	1	1036	99	96
KSR	H	53	119	Protein Kinase	TKL	RAF	1	901	1	873	88	91
KSR2	H	54	120	Protein Kinase	TKL	RAF	51	982	34	849	46	51
KIAA1646	H	55	121	Lipid Kinase	DAG kin	DAG kin	57	537	1	481	89	100
DGK-beta	H	56	122	Lipid Kinase	DAG kin	DAG kin	1	804	1	804	100	100
IP6K1	H	57	123	Lipid Kinase	Inositol kinase	IP6K	1	441	22	462	100	95
YABI	H	58	124	Atypical PK	Atypical	ABC1	280	647	1	368	56	100
AF052122	H	59	125	Atypical PK	Atypical	ABC1	206	591	1	386	65	99
AAF23326	H	60	126	Atypical PK	Atypical	ABC1	1	455	1	455	100	100
SGK493	H	61	127	Atypical PK	Atypical	RIO1	1	552	1	552	100	100
BRD2	H	62	128	Atypical PK	BRD	BRD	1	801	1	801	100	100
BRD3	H	63	129	Atypical PK	BRD	BRD	1	726	1	726	100	100
BRD4	H	64	130	Atypical PK	BRD	BRD	1	722	1	722	100	100
BRDT	H	65	131	Atypical PK	BRD	BRD	1	947	1	947	100	100
ZC1	H	66	132	Protein Kinase	STE	STE20	1	1392	1	1233	87	98

257

TABLE 3
Single Nucleotide Polymorphisms

Gene	ID#na	ID#aa		Nucleotide #	Polymorphism	Nucleotide in patent sequence	AA Residue #	Silent / Residue Change	AA Residue in Patent	Accession#
CRIK	1	67		7676	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1631920_allelePos=201
CRIK	1	67		2933	Y (C/T)	T	961	E / A	A	gnl|dbSNP|ss1337341_allelePos=267
CRIK	1	67		2924	R (A/G)	A	958	silent	T	gnl|dbSNP|ss1337340_allelePos=258
CRIK	1	67		3377	R (A/G)	A	1109	silent	R	gnl|dbSNP|ss1631893_allelePos=310
CRIK	1	67		4085	Y (C/T)	C	1345	silent	S	gnl|dbSNP|ss1631686_allelePos=605
DMPK2	2	68		5050	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1752530_allelePos=201
DMPK2	2	68		1139	R (A/G)	G	358	silent	G	gnl|dbSNP|ss1754079_allelePos=201
MAST3	3	69		2900	Y (C/T)	C	955	silent	D	gnl|dbSNP|ss1846926_allelePos=432
MAST3	3	69		623	Y (C/T)	C	196	silent	H	gnl|dbSNP|ss88979_allelePos=57
MAST205	4	70		2739	R (A/G)	A	913	silent	S	gnl|dbSNP|ss1363030_allelePos=144
MAST205	4	70		25	Y (C/T)	C	9	R / stop	R	gnl|dbSNP|ss133576_allelePos=22
MAST205	4	70		5303	Y (C/T)	C	1768	S / F	S	gnl|dbSNP|ss1529170_allelePos=51
MAST205	4	70		4652	R (A/G)	A	1551	D/G	D	gnl|dbSNP|ss1529101_allelePos=5
MAST205	4	70		3590	R (A/G)	A	1197	K/R	K	gnl|dbSNP|ss1529096_allelePos=51
MAST205	4	70		156	R (A/G)	G	52	silent	A	gnl|dbSNP|ss1608593_allelePos=756
MASTL	5	71		162	S (C/G)	C	54	silent	P	gnl|dbSNP|ss497486_allelePos=201
PKC_eta	6	72		3831	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1363_allelePos=40
PKC_eta	6	72		1840	Y (C/T)	T	558	silent	N	gnl|dbSNP|ss1000395_allelePos=101
PKC_eta	6	72		1239	Y (C/T)	T	358	T/I	I	gnl|dbSNP|ss1472906_allelePos=327
PKC_eta	6	72		2288	S (C/G)	C	3' UTR	-	-	gnl|dbSNP|ss1548761_allelePos=51
H19102	6	72		681	R (A/G)	A	172	H/G	H	gnl|dbSNP|ss1509877_allelePos=51
H19102	7	73		None	-	-	-	-	-	
MSK1	8	74		3186	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss2025310_allelePos=201
MSK1	8	74		3658	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1530678_allelePos=5
MSK1	8	74		3769	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1530679_allelePos=51
MSK1	8	74		3432	K (G/T)	T	3' UTR	-	-	gnl|dbSNP|ss1530677_allelePos=51
MSK1	8	74		3779	K (G/T)	T	3' UTR	-	-	gnl|dbSNP|ss1530680_allelePos=51
YANK3	9	75		1852	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss18125_allelePos=101
YANK3	9	75		1895	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1517863_allelePos=5
YANK3	9	75		2021	M (A/C)	A	3' UTR	-	-	gnl|dbSNP|ss1517888_allelePos=51
MARK2	10	76		2570	Y (C/T)	C	724	silent	S	gnl|dbSNP|ss1121403_allelePos=101
MARK2	10	76		2615	R (A/G)	G	739	silent	P	gnl|dbSNP|ss1121404_allelePos=101
MARK2	10	76		1641	S (C/G)	G	415	P/A	A	gnl|dbSNP|ss1537647_allelePos=51
MARK2	10	76		1547	R (A/G)	A	383	silent	L	gnl|dbSNP|rs1057176_allelePos=51
NuaK2	11	77		1670	S (C/G)	G	538	silent	L	gnl|dbSNP|ss1295001_allelePos=93
NuaK2	11	77		1727	R (A/G)	G	557	silent	L	gnl|dbSNP|ss1295000_allelePos=36
BRSK2	12	78		None	-					
MARK4	13	79		2916	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1967699_allelePos=201
MARK4	13	79		3032	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1967700_allelePos=242
MARK4	13	79		1699	Y (C/T)	C	561	silent	R	gnl|dbSNP|ss1967693_allelePos=201
MARK4	13	79		3092	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1512875_allelePos=51
DCAMKL2	14	80		None	-	-	-	-	-	-
PIM2	15	81		630	R (A/G)	A	210	silent	E	gnl|dbSNP|ss1525746_allelePos=5
PIM2	15	81		1749	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1525747_allelePos=5
PIM2	15	81		1990	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1525754_allelePos=51
PIM3	16	82		2057	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1548948_allelePos=51
PIM3	16	82		1269	Y (C/T)	C	278	silent	P	gnl|dbSNP|ss1511148_allelePos=51
PIM3	16	82		2362	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1511284_allelePos=51
TSSK4	17	83		1203	R (A/G)	A	196	Q/R	Q	gnl|dbSNP|ss1975997_allelePos=201
TSSK4	17	83		152	M (A/C)	C	5' UTR	-	-	gnl|dbSNP|ss1588747_allelePos=749
TSSK4	17	83		141	R (A/G)	A	5' UTR	-	-	gnl|dbSNP|ss1588746_allelePos=738
TSSK4	17	83		238	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss1211997_allelePos=524
TSSK4	17	83		84	Y (C/T)	T	5' UTR	-	-	gnl|dbSNP|ss934800_allelePos=307
TSSK4	17	83		281	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss1747635_allelePos=2506
TSSK4	17	83		236	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1747634_allelePos=2461
TSSK4	17	83		136	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss2056655_allelePos=355
TSSK4	17	83		22	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss45790_allelePos=479
TSSK4	17	83		243	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss2061784_allelePos=1157
TSSK4	17	83		226	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss2061783_allelePos=1140
TSSK4	17	83		47	R (A/G)	A	5' UTR	-	-	gnl|dbSNP|ss1990388_allelePos=1229
TSSK4	17	83		158	W (A/T)	A	5' UTR	-	-	gnl|dbSNP|ss1911350_allelePos=370
TSSK4	17	83		77	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1909793_allelePos=506
TSSK4	17	83		137	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss1909525_allelePos=1475
TSSK4	17	83		44	Y (C/T)	T	5' UTR	-	-	gnl|dbSNP|ss1897673_allelePos=1677
TSSK4	17	83		11	R (A/G)	A	5' UTR	-	-	gnl|dbSNP|ss1857878_allelePos=1145
TSSK4	17	83		223	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1815570_allelePos=267
TSSK4	17	83		85	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss1799649_allelePos=306
TSSK4	17	83		280	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1732367_allelePos=496
TSSK4	17	83		97	Y (C/T)	T	5' UTR	-	-	gnl|dbSNP|ss1729218_allelePos=408
TSSK4	17	83		148	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1684407_allelePos=417
CKIL2	18	84		3889	S (C/G)	C	1208	H / D	H	gnl|dbSNP|ss1551913_allelePos=51
PCTAIRE3	19	85		None	-	-	-	-	-	
PFTAIRE2	20	86		None	-	-	-	-	-	
ERK7	21	87		None	-	-	-	-	-	
CKIlar	22	88		1103	M (A/C)	C	318	silent	A	gnl|dbSNP|ss1537202_allelePos=51
CKIlar	22	88		1008	M (A/C)	C	267	S / R	R	gnl|dbSNP|ss1537192_allelePos=51
CKIlar	22	88		663	Y (C/T)	C	172	R / stop	R	gnl|dbSNP|ss1537165_allelePos=51
CKIlar	22	88		1428	M (A/C)	A	3' UTR	-	-	gnl|dbSNP|ss1537238_allelePos=51

258

TABLE 3 Cont'd
Single Nucleotide Polymorphisms

Gene	ID#na	ID#aa		Nucleotide #	Polymorphism	Nucleotide in patent sequence	AA Residue #	Silent / Residue Change	AA Residue in Patent	Accession#
CKIIar	22	88		194	Y (C/T)	T	15	silent	V	gnl|dbSNP|ss5453_allelePos=51
CKIIar	22	88		1200	R (A/G)	G	351	M / V	V	gnl|dbSNP|ss1537218_allelePos=5
CKIIar	22	88		1181	R (A/G)	A	344	silent	T	gnl|dbSNP|ss1537218_allelePos=51
CKIIar	22	88		1104	W (A/T)	A	319	M / L	M	gnl|dbSNP|ss1537203_allelePos=51
DYRK4	23	89		269	R (A/G)	G	90	R / H	R	gnl|dbSNP|ss88136_allelePos=155
HIPK1	24	90		4114	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss12250_allelePos=101
HIPK4	25	91		None						
BIKE	26	92		1606	R (A/G)	A	468	silent	Q	gnl|dbSNP|ss1509438_allelePos=51
NEK10	27	93		1149	S (C/G)	G	325	T/S	S	gnl|dbSNP|ss727804_allelePos=20
NEK10	27	93		1849	R (A/G)	G	558	silent	G	gnl|dbSNP|ss1691242_allelePos=201
NEK10	27	93		2967	R (A/G)	G	931	N/S	S	gnl|dbSNP|ss1325417_allelePos=338
pNEK5	28	94		None	-	-	-	-	-	-
NEK1	29	95		5063	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1520330_allelePos=51
NEK1	29	95		4848	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1520329_allelePos=51
NEK3	30	96		1854	S (C/G)	C	3' UTR	-	-	gnl|dbSNP|ss3403_allelePos=2
SGK069	31	97		1001	S (C/G)	G	298	P / A	A	gnl|dbSNP|ss1317629_allelePos=393
SGK069	31	97		323	Y (C/T)	C	72	R / C	R	gnl|dbSNP|ss1688815_allelePos=201
SGK110	32	98		299	W (A/T)	A	1	M / L	M	gnl|dbSNP|ss767141_allelePos=201
SGK110	32	98		985	R (A/G)	A	229	silent	P	gnl|dbSNP|ss827468_allelePos=20
SGK110	32	98		640	Y (C/T)	C	114	silent	L	gnl|dbSNP|ss661406_allelePos=201
NRBP2/SGK034	33	99		None	-	-	-	-	-	-
CNK	34	100		None	-	-	-	-	-	-
SCYL2/AI05225	35	101		None	-	-	-	-	-	-
SRPK2	36	102		2219	Y (C/T)	C	681	L / F	L	gnl|dbSNP|ss1525084_allelePos=51
SRPK2	36	102		2047	Y (C/T)	C	623	silent	F	gnl|dbSNP|ss1525076_allelePos=51
SRPK2	36	102		2040	R (A/G)	G	621	Q / R	R	gnl|dbSNP|ss1525074_allelePos=51
SRPK2	36	102		2035	Y (C/T)	T	619	silent	Y	gnl|dbSNP|rs1050422_allelePos=51
SRPK2	36	102		2021	M (A/C)	C	615	I / L	L	gnl|dbSNP|ss1525069_allelePos=51
SRPK2	36	102		2014	M (A/C)	C	612	Q / H	H	gnl|dbSNP|ss1525068_allelePos=51
SRPK2	36	102		2029	R (A/G)	A	617	silent	G	gnl|dbSNP|ss1525072_allelePos=51
SRPK2	36	102		2017	Y (C/T)	T	613	silent	F	gnl|dbSNP|ss1525068_allelePos=51
SRPK2	36	102		2016	W (A/T)	T	613	Y / F	F	gnl|dbSNP|ss1525067_allelePos=51
SRPK2	36	102		2001	R (A/G)	G	608	N / S	S	gnl|dbSNP|ss1525064_allelePos=51
SRPK2	36	102		1999	S (C/G)	C	607	silent	G	gnl|dbSNP|ss1525063_allelePos=51
SRPK2	36	102		1996	R (A/G)	A	606	silent	A	gnl|dbSNP|ss1525062_allelePos=51
SRPK2	36	102		1969	Y (C/T)	C	597	silent	D	gnl|dbSNP|ss1525061_allelePos=51
SRPK2	36	102		2044	R (A/G)	A	622	silent	E	gnl|dbSNP|ss1525075_allelePos=51
SRPK2	36	102		2023	R (A/G)	A	615	silent	L	gnl|dbSNP|ss1525072_allelePos=51
TLK1	37	103		2174	W (A/T)	A	646	V/D	D	gnl|dbSNP|ss1515391_allelePos=51
TLK1	37	103		2489	R (A/G)	A	751	N/S	N	gnl|dbSNP|ss1515399_allelePos=51
TLK1	37	103		2515	M (A/C)	A	760	silent	R	gnl|dbSNP|ss1515400_allelePos=51
TLK1	37	103		2358	R (A/G)	A	707	silent	E	gnl|dbSNP|ss1515395_allelePos=51
TLK1	37	103		2294	W (A/T)	T	686	Y / F	F	gnl|dbSNP|ss1515394_allelePos=51
TLK1	37	103		2229	R (A/G)	A	664	silent	V	gnl|dbSNP|ss1515393_allelePos=51
TLK1	37	103		2014	Y (C/T)	C	593	silent	L	gnl|dbSNP|ss1515384_allelePos=51
TLK1	37	103		1137	W (A/T)	T	300	silent	I	gnl|dbSNP|ss1515380_allelePos=51
TLK1	37	103		3279	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1515413_allelePos=51
TLK1	37	103		3142	S (C/G)	G	3' UTR	-	-	gnl|dbSNP|ss1515412_allelePos=51
TLK1	37	103		2488	W (A/T)	A	751	N / Y	N	gnl|dbSNP|ss1515396_allelePos=51
TLK1	37	103		1711	K (G/T)	T	492	D/Y	Y	gnl|dbSNP|ss1515382_allelePos=51
TLK1	37	103		1730	M (A/C)	A	498	S/Y	Y	gnl|dbSNP|ss1515383_allelePos=51
TLK1	37	103		1083	R (A/G)	A	282	E/D	E	gnl|dbSNP|ss1515377_allelePos=51
TLK1	37	103		1647	Y (C/T)	C	470	silent	H	gnl|dbSNP|ss1515381_allelePos=51
TLK1	37	103		1092	R (A/G)	A	285	silent	K	gnl|dbSNP|ss1515379_allelePos=51
TLK1	37	103		1035	Y (C/T)	T	266	silent	A	gnl|dbSNP|ss1515376_allelePos=51
TLK1	37	103		951	R (A/G)	A	238	silent	T	gnl|dbSNP|ss1515375_allelePos=51
SGK071	38	104		None	-	-	-	-	-	-
SK516	39	105		None	-	-	-	-	-	-
H85389	40	106		None	-	-	-	-	-	-
Wee1b/SGK461	41	107		None	-	-	-	-	-	-
Wnk2	42	108		7079	K (G/T)	T	3' UTR	-	-	gnl|dbSNP|ss2899_allelePos=78
MAP3K1	43	109		2716	R (A/G)	A	906	I / V	I	gnl|dbSNP|ss1317910_allelePos=285
MAP3K1	43	109		6227	W (A/T)	A	3' UTR	-	-	gnl|dbSNP|ss1146242_allelePos=109
MAP3K1	43	109		5560	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss1286358_allelePos=101
MAP3K1	43	109		3187	M (A/C)	C	1063	silent	R	gnl|dbSNP|ss1146312_allelePos=101
MAP3K1	43	109		6015	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1146243_allelePos=101
MAP3K1	43	109		2416	R (A/G)	A	806	N / D	N	gnl|dbSNP|ss1146310_allelePos=101
MAP3K1	43	109		1284	R (A/G)	A	428	silent	T	gnl|dbSNP|ss1146300_allelePos=101
MAP3K8	44	110		247	S (C/G)	G	83	Q/E	E	gnl|dbSNP|ss1394913_allelePos=101
MAP3K8	44	110		2485	K (G/T)	T	3' UTR	-	-	gnl|dbSNP|ss1817_allelePos=49
MAP3K8	44	110		2298	M (A/C)	A	3' UTR	-	-	gnl|dbSNP|ss1547718_allelePos=51
Pak4_m	45	111		None	-	-	-	-	-	-
STLK6r	46	112		487	R (A/G)	G	82	silent	T	gnl|dbSNP|ss1483412_allelePos=100
Map2K2	47	113		904	M (A/C)	C	219	silent	I	gnl|dbSNP|ss1937135_allelePos=201
CCK4	48	114		3636	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1527472_allelePos=51
LMR1	49	115		None	-	-	-	-	-	-
RYK	50	116		2875	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss16914_allelePos=101
RYK	50	116		2496	W (A/T)	A	3' UTR	-	-	gnl|dbSNP|ss1525573_allelePos=51

TABLE 3 Cont'd
Single Nucleotide Polymorphisms

Gene	ID#na	ID#aa		Nucleotide #	Polymorphism	Nucleotide in patent sequence	AA Residue #	Silent / Residue Change	AA Residue in Patent	Accession#
RYK	50	116		851	R (A/G)	G	254	N / S	S	gnl|dbSNP|ss1525514_allelePos=51
RYK	50	116		386	R (A/G)	G	99	N / S	S	gnl|dbSNP|ss1525513_allelePos=51
RYK	50	116		2764	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss16913_allelePos=31
LRRK2	51	117		5425	W (A/T)	T	1598	E/V	V	gnl|dbSNP|ss63276_allelePos=97
pMLK4	52	118		3597	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss2057123_allelePos=323
pMLK4	52	118		3914	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss2057120_allelePos=201
pMLK4	52	118		3668	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss2057122_allelePos=288
pMLK4	52	118		3800	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss2057121_allelePos=22
pMLK4	52	118		2580	Y (C/T)	C	773	silent	S	gnl|dbSNP|ss1411720_allelePos=519
pMLK4	52	118		2611	K (G/T)	T	784	G/C	C	gnl|dbSNP|ss1411719_allelePos=488
pMLK4	52	118		4193	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss2057119_allelePos=201
pMLK4	52	118		4309	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss2057118_allelePos=201
KSR	53	119		4096	S (C/G)	C	3' UTR	-	-	gnl|dbSNP|ss100899_allelePos=172
KSR2	54	120		612	S (C/G)	C	204	silent	T	gnl|dbSNP|ss2005786_allelePos=201
KIAA1646	55	121		3769	M (A/C)	A	3' UTR	-	-	gnl|dbSNP|ss2052346_allelePos=499
KIAA1646	55	121		3020	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss2052345_allelePos=201
KIAA1646	55	121		2577	K (G/T)	T	3' UTR	-	-	gnl|dbSNP|ss2052344_allelePos=201
KIAA1646	55	121		2391	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss2052344_allelePos=201
KIAA1646	55	121		4272	R (A/G)	A	3' UTR	-	-	gnl|dbSNP|ss2052347_allelePos=201
DGK-beta	56	122		None	-	-	-	-	-	
IP6K1	57	123		3669	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1522850_allelePos=51
IP6K1	57	123		2851	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1522846_allelePos=51
YAB1	58	124		2506	R (A/G)	G	3' UTR	-	-	gnl|dbSNP|ss1305707_allelePos=99
YAB1	58	124		1538	Y (C/T)	C	480	silent	F	gnl|dbSNP|ss1529336_allelePos=51
AF052122	59	125		None	-	-	-	-	-	
AAF23326	60	126		None	-	-	-	-	-	
SGK493	61	127		1094	R (A/G)	A	349	R / G	R	gnl|dbSNP|ss1826551_allelePos=201
SGK493	61	127		1690	Y (C/T)	T	547	silent	A	gnl|dbSNP|ss1826528_allelePos=201
BRD2	62	128		920	K (G/T)	T	5' UTR	-	-	gnl|dbSNP|ss1425392_allelePos=324
BRD2	62	128		1794	R (A/G)	A	31	silent	K	gnl|dbSNP|ss686785_allelePos=201
BRD2	62	128		3510	Y (C/T)	T	603	silent	S	gnl|dbSNP|rs516535_allelePos=201
BRD2	62	128		2413	Y (C/T)	C	238	L / F	L	gnl|dbSNP|ss1973307_allelePos=201
BRD2	62	128		3199	K (G/T)	G	500	E/stop	E	gnl|dbSNP|ss15121_allelePos=101
BRD2	62	128		3333	R (A/G)	G	544	silent	K	gnl|dbSNP|ss13218_allelePos=101
BRD2	62	128		4348	M (A/C)	C	3' UTR	3' UTR	-	gnl|dbSNP|ss12998_allelePos=101
BRD2	62	128		3411	Y (C/T)	T	570	silent	D	gnl|dbSNP|ss1550506_allelePos=51
BRD2	62	128		1344	R (A/G)	G	5' UTR	-	-	gnl|dbSNP|ss1550448_allelePos=51
BRD2	62	128		4416	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss1550446_allelePos=51
BRD2	62	128		4219	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1523158_allelePos=51
BRD2	62	128		3342	R (A/G)	G	547	silent	R	gnl|dbSNP|ss1523069_allelePos=51
BRD2	62	128		811	Y (C/T)	C	5' UTR	-	-	gnl|dbSNP|ss1522874_allelePos=51
BRD2	62	128		2379	S (C/G)	G	226	silent	L	gnl|dbSNP|ss16333_allelePos=31
BRD3	63	129		2405	Y (C/T)	T	3' UTR	-	-	gnl|dbSNP|ss575919_allelePos=201
BRD3	63	129		1075	R (A/G)	G	312	silent	L	gnl|dbSNP|ss630265_allelePos=201
BRD3	63	129		1975	Y (C/T)	C	612	silent	D	gnl|dbSNP|ss901346_allelePos=201
BRD3	63	129		1423	Y (C/T)	C	428	silent	P	gnl|dbSNP|ss834964_allelePos=201
BRD3	63	129		2934	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss17101_allelePos=101
BRD3	63	129		2796	Y (C/T)	C	3' UTR	-	-	gnl|dbSNP|ss1527035_allelePos=51
BRD4	64	130		1846	R (A/G)	G	542	N / D	D	gnl|dbSNP|ss1512910_allelePos=51
BRDT	65	131		821	M (A/C)	A	238	K / N	K	gnl|dbSNP|ss1559581_allelePos=482
BRDT	65	131		2976	M (A/C)	C	3' UTR	-	-	gnl|dbSNP|ss1553268_allelePos=51
BRDT	65	131		2785	M (A/C)	C	893	Q / P	P	gnl|dbSNP|ss1553264_allelePos=51
BRDT	65	131		1114	M (A/C)	C	336	stop / S	S	gnl|dbSNP|ss1553262_allelePos=51
BRDT	65	131		1113	W (A/T)	T	336	Y/S	S	gnl|dbSNP|ss1553261_allelePos=51
BRDT	65	131		2882	M (A/C)	C	925	silent	A	gnl|dbSNP|ss1553267_allelePos=51
BRDT	65	131		2851	M (A/C)	C	915	Q/P	P	gnl|dbSNP|ss1553266_allelePos=51
BRDT	65	131		2846	M (A/C)	C	913	silent	A	gnl|dbSNP|ss1553265_allelePos=51
ZC1	66	132		1382	R (A/G)	A	418	silent	E	gnl|dbSNP|rs1139583_allelePos=51
ZC1	66	132		2684	S (C/G)	G	852	silent	S	gnl|dbSNP|rs1042916_allelePos=51

TABLE 4
Protein Domains

Gene	ID#na	ID#aa	Profile Description	Profile Accession	Pscore	Domain Start	Domain End	Profile Start	Profile End	Profile Length	Query Length
CRIK	1	67	Protein Kinase domain	PF00069	9.20E-67	98	361	1	278	278	2055
CRIK	1	67	CNH domain	PF00780	2.60E-115	1620	1917	1	378	378	2055
CRIK	1	67	PH domain	PF00169	3.00E-16	1472	1591	1	85	85	2055
CRIK	1	67	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	1.00E-09	1391	1439	1	51	51	2055
CRIK	1	67	Protein kinase C terminal domain	PF00433	3.00E-08	382	391	1	32	70	2055
DMPK2	2	68	Protein kinase domain	PF00069	2.10E-70	71	337	1	278	278	1572
DMPK2	2	68	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	3.10E-17	887	935	1	51	51	1572
DMPK2	2	68	PH domain	PF00169	1.70E-16	956	1074	1	85	85	1572
DMPK2	2	68	CNH domain	PF00780	1.50E-12	1100	1380	1	378	378	1572
DMPK2	2	68	Protein kinase C terminal domain	PF00433	2.00E-08	351	366	16	31	70	1572
MAST3	3	69	Protein Kinase domain	PF00069	5.50E-74	389	635	1	149	294	1331
MAST3	3	69	Protein Kinase domain	PF00069	5.50E-74	580	662	158	294	294	1331
MAST3	3	69	PDZ doamin	PF00595	3.70E-09	972	1054	1	79	84	1331
MAST205	4	70	Protein kinase domain	PF00069	7.90E-80	512	785	1	278	278	1798
MAST205	4	70	PDZ domain (Also known as DHR or GLGF)	PF00595	2.20E-10	1104	1191	1	83	83	1798
MASTL	5	71	Protein kinase domain	PF00069	2.20E-73	35	310	1	278	278	878
MASTL	5	71	Protein kinase domain	PF00069	2.20E-73	739	834	149	278	278	878
MASTL	5	71	Protein kinase C terminal domain	PF00433	4.60E-07	835	883	1	31	70	878
PKC_eta	6	72	Protein kinase domain	PF00069	3.60E-82	355	614	1	294	294	683
PKC_eta	6	72	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	4.40E-46	172	222	1	51	51	683
PKC_eta	6	72	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	4.40E-46	246	295	1	51	51	683
PKC_eta	6	72	Protein kinase C terminal domain	PF00433	1.80E-41	615	681	1	70	70	683
H19102	7	73	Protein kinase domain	PF00069	3.20E-64	146	398	1	278	278	449
MSK1	8	74	Protein kinase domain	PF00069	1.60E-182	49	318	1	278	278	802
MSK1	8	74	Protein kinase domain	PF00069	1.60E-182	427	687	2	278	278	802
MSK1	8	74	Protein kinase C terminal domain	PF00433	2.40E-21	319	382	1	70	70	802
YANK3	9	75	Protein kinase domain	PF00069	3.80E-71	93	345	1	287	294	486
MARK2	10	76	Protein kinase domain	PF00069	1.30E-100	53	304	1	294	294	787
MARK2	10	76	Kinase associated domain 1	PF02149	3.00E-21	738	787	1	50	60	787
MARK2	10	76	UBA/TS-N domain	PF00627	0.000003	324	363	1	45	45	787
NuaK2	11	77	Protein kinase domain	PF00069	8.00E-94	97	347	1	294	294	672
BRSK2	12	78	Protein kinase domain	PF00069	3.20E-97	19	270	1	278	278	674
MARK4	13	79	Protein kinase domain	PF00069	7.70E-104	59	310	1	278	278	752
MARK4	13	79	Kinase associated domain 1	PF02149	1.30E-15	703	752	1	50	50	752
MARK4	13	79	UBA domain	PF00627	6.30E-11	330	368	1	41	41	752
DCAMKL2	14	80	Protein kinase domain	PF00069	1.70E-97	394	651	1	278	278	766
PIM2	15	81	Protein kinase domain	PF00069	1.40E-71	132	388	1	294	294	434
PIM3	16	82	Protein kinase domain	PF00069	9.90E-80	40	293	1	278	278	326

261

TABLE 4 Cont'd
Protein Domains

Gene	ID#/na	ID#/aa	Profile Description	Profile Accession	Pscore	Domain Start	Domain End	Profile Start	Profile End	Profile Length	Query Length
TSSK4	17	83	Protein kinase domain	PF00069	1.10E-78	25	293	1	278	278	328
CKIL2	18	84	Protein kinase domain	PF00069	8.50E-33	21	276	1	265	278	1244
PCTAIRE3	19	85	Protein kinase domain	PF00069	1.20E-87	50	331	1	278	278	380
PFTAIRE2	20	86	Protein kinase domain	PF00069	4.40E-80	103	387	1	278	278	435
ERK7	21	87	Protein kinase domain	PF00069	4.80E-90	13	323	1	278	278	563
CKIIa-rs	22	88	Protein kinase domain	PF00069	2.20E-89	39	324	1	278	278	391
DYRK4	23	89	Protein kinase domain	PF00069	4.00E-84	506	802	1	278	278	921
HIPK1	24	90	Protein kinase domain	PF00069	6.20E-58	190	518	1	278	278	1210
HIPK4	25	91	Protein kinase domain	PF00069	1.10E-58	11	347	1	278	278	616
BIKE	26	92	Protein kinase domain	PF00069	2.50E-38	51	314	1	294	294	1161
NEK10	27	93	Protein kinase domain	PF00069	8.80E-70	519	783	1	294	294	1125
NEK10	27	93	Armadillo/beta-catenin-like repeat	PF00514	0.009707	198	238	1	40	40	1125
NEK10	27	93	Armadillo/beta-catenin-like repeat	PF00514	0.009707	239	279	1	40	40	1125
NEK10	27	93	Armadillo/beta-catenin-like repeat	PF00514	0.009707	280	320	1	40	40	1125
pNEK5	28	94	Protein kinase domain	PF00069	9.10E-87	61	316	1	294	294	889
NEK1	29	95	Protein kinase domain	PF00069	2.50E-89	4	258	1	278	278	1286
NEK3	30	96	Protein kinase domain	PF00069	5.60E-92	4	257	1	278	278	506
SGK069	31	97	Protein kinase domain	PF00069	3.80E-40	62	325	1	263	278	348
SGK110	32	98	Protein kinase domain	PF00069	1.70E-39	98	359	27	273	278	414
NRBP2	33	99	Protein kinase domain	PF00069	2.00E-24	38	313	1	278	278	507
CNK	34	100	Protein kinase domain	PF00069	1.60E-91	62	314	1	278	278	646
CNK	34	100	POLO box duplicated region.	PF00659	9.70E-35	470	533	1	77	77	646
CNK	34	100	POLO box duplicated region.	PF00659	9.70E-35	567	637	1	77	77	646
SCYL2	35	101	Protein kinase domain	PF00069	8.00E-13	32	327	1	278	278	933
SRPK2	36	102	Protein kinase domain	PF00069	7.40E-42	81	686	1	278	278	688
TLK1	37	103	Protein kinase domain	PF00069	4.70E-71	477	755	1	278	278	787
SGK071	38	104	Protein kinase domain	PF00069	7.60E-26	28	296	27	278	278	632
SK618	39	105	Protein kinase domain	PF00069	2.50E-44	652	915	1	278	278	929
H85389	40	106	Protein kinase domain	PF00069	3.90E-60	69	397	1	278	278	401
Wee1b	41	107	Protein kinase domain	PF00069	1.10E-49	212	488	1	272	278	557

TABLE 4 Cont'd
Protein Domains

Gene	ID#na	ID#aa	Profile Description	Profile Accession	Pscore	Domain Start	Domain End	Profile Start	Profile End	Profile Length	Query Length
Wnk2	42	108	Protein kinase domain	PF00069	6.60E-63	181	439	1	278	278	2245
MAP3K1	43	109	Protein kinase domain	PF00069	1.00E-85	1242	1507	1	278	278	1511
MAP3K8	44	110	Protein kinase domain	PF00069	2.10E-88	468	731	1	278	278	735
Pak4	45	111	Protein kinase domain	PF00069	5.00E-86	323	574	1	278	278	593
Pak4	45	111	P21-Rho-binding domain	PF00786	3.20E-12	11	69	1	64	64	593
STLK6-rs	46	112	Protein kinase domain	PF00069	2.60E-33	58	369	14	278	278	418
MAP2K2	47	113	Protein kinase domain	PF00069	3.20E-58	72	369	1	278	278	381
CCK4	48	114	Protein kinase domain	PF00069	6.70E-63	796	1061	1	272	278	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	46	103	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	143	202	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	239	303	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	336	393	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	426	483	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	517	572	1	45	45	1070
CCK4	48	114	Immunoglobulin domain	PF00047	1.00E-61	606	666	1	45	45	1070
LMR1	49	115	Protein kinase domain	PF00069	1.10E-46	125	395	1	294	294	1374
RYK	50	116	Protein kinase domain	PF00069	3.10E-81	330	596	1	276	278	607
RYK	50	115	WIF domain	PF02019	3.30E-91	66	194	1	132	132	607
LRRK2	51	117	Protein kinase domain	PF00069	1.00E-41	1866	2138	8	272	278	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	983	1004	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1012	1035	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1036	1058	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1084	1103	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1108	1129	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1130	1153	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1174	1196	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1197	1218	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1221	1244	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1246	1268	1	23	23	2534
LRRK2	51	117	Leucine Rich Repeat	PF00560	2.10E-34	1269	1293	1	23	23	2534
pMLK4	52	118	Protein kinase domain	PF00069	1.70E-87	124	398	1	292	294	1036
pMLK4	52	118	SH3 domain	PF00018	2.00E-14	45	100	5	58	58	1036
KSR	53	119	Protein kinase domain	PF00069	1.40E-31	591	731	1	147	294	901
KSR	53	119	Protein kinase domain	PF00069	1.40E-31	753	792	163	195	294	901
KSR	53	119	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	0.008623	348	391	1	51	51	901
KSR	53	119	MYND finger	PF01753	1.311685	360	377	1	21	43	901
KSR2	54	120	Protein kinase domain	PF00069	6.90E-40	698	957	1	289	294	982
KSR2	54	120	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	0.000127	445	488	1	51	51	982
KIAA1646	55	121	Diacylglycerol kinase catalytic domain	PF00781	2.50E-09	132	278	1	159	159	537

TABLE 4 Cont'd
Protein Domains

Gene	ID#na	ID#aa	Profile Description	Profile Accession	Pscore	Domain Start	Domain End	Profile Start	Profile End	Profile Length	Query Length
DGK-beta	56	122	Diacylglycerol kinase accessory domain	PF00609	3.30E-129	582	762	1	180	180	804
DGK-beta	56	122	Diacylglycerol kinase catalytic domain	PF00781	1.20E-71	438	562	1	159	159	804
DGK-beta	56	122	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	5.00E-28	245	294	1	51	51	804
DGK-beta	56	122	Phorbol esters/diacylglycerol binding domain (C1 domain)	PF00130	5.00E-28	310	358	1	51	51	804
DGK-beta	56	122	EF hand	PF00036	4.10E-17	153	181	1	29	29	804
DGK-beta	56	122	EF hand	PF00036	4.10E-17	188	226	1	29	29	804
IP6K1	57	123	No domain identified								
YAB1	58	124	ABC1 family	PF03109	1.20E-42	318	434	1	124	124	647
AF052122	59	125	No domain identified								
AAF23326	60	126	No domain identified								
SGK493	61	127	No domain identified								
BRD2	62	128	Bromodomain	PF00439	4.80E-91	79	168	1	92	92	801
BRD2	62	128	Bromodomain	PF00439	4.90E-91	352	441	1	92	92	801
BRD3	63	129	Bromodomain	PF00439	6.50E-87	39	128	1	92	92	726
BRD3	63	129	Bromodomain	PF00439	6.50E-87	315	403	1	92	92	726
BRD4	64	130	Bromodomain	PF00439	1.80E-90	63	152	1	92	92	722
BRD4	64	130	Bromodomain	PF00439	1.80E-90	356	445	1	92	92	722
BRDT	65	131	Bromodomain	PF00439	7.50E-86	32	121	1	92	92	947
BRDT	65	131	Bromodomain	PF00439	7.50E-86	275	384	1	92	92	947
ZC1	66	132	CNH	PF00780	9.20E-131	1066	1372	1	378	378	1392
ZC1	66	132	Protein kinase domain	PF00069	1.40E-91	25	289	1	278	278	1392

TABLE 5
Chromosomal Data

Gene_NAME	Sp	ID#na	ID#aa	Cytogenetic position	Cancer Amplicon	Disease Loci
CRIK	H	1	67	12q24.31		
DMPK2	H	2	68	11q12-q13.1	11q13-q14	Osteoarthritis OMIM 165720
MAST3	H	3	69	19p13.1		
MAST205	H	4	70	1p34.1		
MASTL	H	5	71	10p11.2-p12.1		Schizophrenia, OMIM 181500
PKC_eta	H	6	72	14q23.1		
H19102	H	7	73	17q11.1	17q12-q21	
MSK1	H	8	74	14q32.11		
YANK3	H	9	75	10q26.3		
MARK2	H	10	76	11q12-11q13	11q13-q14	Osteoarthritis OMIM 165720
NuaK2	H	11	77	1q31-q32.1		
BRSK2	H	12	78	11p15.5		
MARK4	H	13	79	19q13.2-q13.33	19cen-q13.3	
DCAMKL2	H	14	80	4q31.3		
PIM2	H	15	81	Xp11.23	Xp11.2-p21	
PIM3	H	16	82	22q13		
TSSK4	H	17	83	14q11.1		
CKIL2	H	18	84	15q14-q15.3		Schizophrenia, 15q15, OMIM 181500
PCTAIRE3	H	19	85	1q32		
PFTAIRE2	H	20	86	2q33.2-q34	2q31-q33	Pulmonary Hypertension, 2q33, OMIM 178600; Osteoarthritis, 2q34-q35, OMIM 140600
ERK7	H	21	87	8q24.3		
CKIIa-rs	H	22	88	11p15		
DYRK4	H	23	89	12p13		Hypertension, essential, 12p13, OMIM 145500
HIPK1	H	24	90	1p11-p12		
HIPK4	H	25	91	19q13.1	19cen-q13.3	Osteoarthritis OMIM 165720
BIKE	H	26	92	4q13-q21.21		
NEK10	H	27	93	3p21.33		
pNEK5	H	28	94	13q14	13q14	
NEK1	H	29	95	4q33-q34		
NEK3	H	30	96	13q14.3	13q14	
SGK069	H	31	97	19q13.43		
SGK110	H	32	98	19q13.43		
NRBP2	H	33	99	8q24.3		
CNK	H	34	100	1p34.1		
SCYL2	H	35	101	12q23-q24.1		
SRPK2	H	36	102	7q22.3	7q21-q22	Osteoarthritis OMIM 165720
TLK1	H	37	103	2q31.1		
SGK071	H	38	104	9q34		
SK516	H	39	105	1q31-32.1		
H85389	H	40	106	20p13		
Wee1b	H	41	107	7q34-36		
Wnk2	H	42	108	9q22.31		
MAP3K1	H	43	109	5q11.2-q13		Schizophrenia, 5q11-q13, OMIM 181500
MAP3K8	H	44	110	2q21.3		
Pak4_m	M	45	111	murine		
STLK6-rs	H	46	112	1p33		
MAP2K2	H	47	113	7q34		
CCK4	H	48	114	6p21-p12		
LMR1	H	49	115	17q25		
RYK	H	50	116	3q22		
LRRK2	H	51	117	12q11-q12		
pMLK4	H	52	118	1q42.2		
KSR	H	53	119	17q11.1	17q12-q21	
KSR2	H	54	120	12q24.3		
KIAA1646	H	55	121	22q13.31		
DGK-beta	H	56	122	7p21.3-p22		Osteoarthritis OMIM 165720
IP6K1	H	57	123	3p21.31		
YAB1	H	58	124	1q42		Schizophrenia, 1q42.1, OMIM 181500
AF052122	H	59	125	19q13.1	19cen-q13.3	
AAF23326	H	60	126	14q24.3-q32		
SGK493	H	61	127	5q14		
BRD2	H	62	128	6p21.2		
BRD3	H	63	129	9q34		
BRD4	H	64	130	19p13.2		
BRDT	H	65	131	1p21		
ZC1	H	66	132	2q11.1-q11.2		

265

TABLE 6
Human ESTs

Rank	Gene	Human EST
1	CRIK_H_SEQID#NA_1	BQ070955.1
2	CRIK_H_SEQID#NA_1	BQ071141.1
3	CRIK_H_SEQID#NA_1	BQ228524.1
4	CRIK_H_SEQID#NA_1	BM545592.1
5	CRIK_H_SEQID#NA_1	BI253509.1
6	CRIK_H_SEQID#NA_1	BG912161.1
7	CRIK_H_SEQID#NA_1	BG252350.1
8	CRIK_H_SEQID#NA_1	BG120427.1
9	CRIK_H_SEQID#NA_1	BE875297.1
10	CRIK_H_SEQID#NA_1	BQ448184.1
1	DMPK2_H_SEQID#NA_2	BI793270.1
2	DMPK2_H_SEQID#NA_2	BI792977.1
3	DMPK2_H_SEQID#NA_2	BG752641.1
4	DMPK2_H_SEQID#NA_2	BG752641.1
5	DMPK2_H_SEQID#NA_2	AW516225.1
6	DMPK2_H_SEQID#NA_2	BG678034.1
7	DMPK2_H_SEQID#NA_2	AA809737.1
8	DMPK2_H_SEQID#NA_2	BE793390.1
9	DMPK2_H_SEQID#NA_2	BE793390.1
10	DMPK2_H_SEQID#NA_2	AW814108.1
1	MAST3_H_SEQID#NA_3	BG765138.1
2	MAST3_H_SEQID#NA_3	BG767919.1
3	MAST3_H_SEQID#NA_3	BF684640.1
4	MAST3_H_SEQID#NA_3	BF346524.1
5	MAST3_H_SEQID#NA_3	BE261265.1
6	MAST3_H_SEQID#NA_3	BF346384.1
7	MAST3_H_SEQID#NA_3	BG257232.1
8	MAST3_H_SEQID#NA_3	BF689544.1
9	MAST3_H_SEQID#NA_3	BI907332.1
10	MAST3_H_SEQID#NA_3	BM966751.1
1	MAST205_H_SEQID#NA_4	BQ231137.1
2	MAST205_H_SEQID#NA_4	BQ070626.1
3	MAST205_H_SEQID#NA_4	AL568230.1
4	MAST205_H_SEQID#NA_4	BQ050660.1
5	MAST205_H_SEQID#NA_4	BM471504.1
6	MAST205_H_SEQID#NA_4	BG831571.1
7	MAST205_H_SEQID#NA_4	AL540100.1
8	MAST205_H_SEQID#NA_4	BI771067.1
9	MAST205_H_SEQID#NA_4	BG762487.1
10	MAST205_H_SEQID#NA_4	BG676428.1
1	MASTL_H_SEQID#NA_5	AL541215.1
2	MASTL_H_SEQID#NA_5	AL520252.1
3	MASTL_H_SEQID#NA_5	BQ441178.1
4	MASTL_H_SEQID#NA_5	BM550518.1
5	MASTL_H_SEQID#NA_5	BQ224736.1
6	MASTL_H_SEQID#NA_5	BM721150.1
7	MASTL_H_SEQID#NA_5	AL712023.1
8	MASTL_H_SEQID#NA_5	BM679574.1
9	MASTL_H_SEQID#NA_5	BG027109.1
10	MASTL_H_SEQID#NA_5	BM748750.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
1	PKC_eta_H_SEQID#NA_6	BM920615.1
2	PKC_eta_H_SEQID#NA_6	BM457208.1
3	PKC_eta_H_SEQID#NA_6	BQ051772.1
4	PKC_eta_H_SEQID#NA_6	AU136862.1
5	PKC_eta_H_SEQID#NA_6	BI913495.1
6	PKC_eta_H_SEQID#NA_6	BG820252.1
7	PKC_eta_H_SEQID#NA_6	BM549890.1
8	PKC_eta_H_SEQID#NA_6	BE161764.1
9	PKC_eta_H_SEQID#NA_6	BG719560.1
10	PKC_eta_H_SEQID#NA_6	BQ006934.1
1	H19102_H_SEQID#NA_7	BI546006.1
2	H19102_H_SEQID#NA_7	BF954472.1
3	H19102_H_SEQID#NA_7	BQ363219.1
4	H19102_H_SEQID#NA_7	H19102.1
5	H19102_H_SEQID#NA_7	BF362477.1
6	H19102_H_SEQID#NA_7	BF362466.1
7	H19102_H_SEQID#NA_7	BF362458.1
8	H19102_H_SEQID#NA_7	AA808745.1
9	H19102_H_SEQID#NA_7	BE968821.1
10	H19102_H_SEQID#NA_7	BE968821.1
1	MSK1_H_SEQID#NA_8	BM556986.1
2	MSK1_H_SEQID#NA_8	BM453259.1
3	MSK1_H_SEQID#NA_8	BG684373.1
4	MSK1_H_SEQID#NA_8	BM968829.1
5	MSK1_H_SEQID#NA_8	BI088037.1
6	MSK1_H_SEQID#NA_8	BE410965.1
7	MSK1_H_SEQID#NA_8	BG699153.1
8	MSK1_H_SEQID#NA_8	AA314565.1
9	MSK1_H_SEQID#NA_8	BM475296.1
10	MSK1_H_SEQID#NA_8	BM690068.1
1	YANK3_H_SEQID#NA_9	BI917132.1
2	YANK3_H_SEQID#NA_9	BI257653.1
3	YANK3_H_SEQID#NA_9	BG824303.1
4	YANK3_H_SEQID#NA_9	BG282899.1
5	YANK3_H_SEQID#NA_9	BM702426.1
6	YANK3_H_SEQID#NA_9	AW245946.1
7	YANK3_H_SEQID#NA_9	AW245503.1
8	YANK3_H_SEQID#NA_9	BG719068.1
9	YANK3_H_SEQID#NA_9	BM666731.1
10	YANK3_H_SEQID#NA_9	BF446773.1
1	MARK2_H_SEQID#NA_10	BM550195.1
2	MARK2_H_SEQID#NA_10	BE795309.1
3	MARK2_H_SEQID#NA_10	BG825423.1
4	MARK2_H_SEQID#NA_10	BE798169.1
5	MARK2_H_SEQID#NA_10	BI521469.1
6	MARK2_H_SEQID#NA_10	AU133733.1
7	MARK2_H_SEQID#NA_10	BE397682.1
8	MARK2_H_SEQID#NA_10	BG822223.1
9	MARK2_H_SEQID#NA_10	BI911013.1
10	MARK2_H_SEQID#NA_10	BE280645.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
1	NuaK2_H_SEQID#NA_11	BM927376.1
2	NuaK2_H_SEQID#NA_11	BQ062868.1
3	NuaK2_H_SEQID#NA_11	BQ064231.1
4	NuaK2_H_SEQID#NA_11	BQ059508.1
5	NuaK2_H_SEQID#NA_11	BQ060729.1
6	NuaK2_H_SEQID#NA_11	BM909401.1
7	NuaK2_H_SEQID#NA_11	BQ056806.1
8	NuaK2_H_SEQID#NA_11	BQ065633.1
9	NuaK2_H_SEQID#NA_11	BQ064127.1
10	NuaK2_H_SEQID#NA_11	BQ056490.1
1	BRSK2_H_SEQID#NA_12	AL538014.1
2	BRSK2_H_SEQID#NA_12	BG395625.1
3	BRSK2_H_SEQID#NA_12	BI825755.1
4	BRSK2_H_SEQID#NA_12	BM677936.1
5	BRSK2_H_SEQID#NA_12	BG395884.1
6	BRSK2_H_SEQID#NA_12	BM805756.1
7	BRSK2_H_SEQID#NA_12	BE251924.1
8	BRSK2_H_SEQID#NA_12	BE550940.1
9	BRSK2_H_SEQID#NA_12	BF525960.1
10	BRSK2_H_SEQID#NA_12	BE259121.1
1	MARK4_H_SEQID#NA_13	BG745114.1
2	MARK4_H_SEQID#NA_13	BM543319.1
3	MARK4_H_SEQID#NA_13	BQ066239.1
4	MARK4_H_SEQID#NA_13	BG389721.1
5	MARK4_H_SEQID#NA_13	BF982422.1
6	MARK4_H_SEQID#NA_13	BM467107.1
7	MARK4_H_SEQID#NA_13	BG744466.1
8	MARK4_H_SEQID#NA_13	BG760697.1
9	MARK4_H_SEQID#NA_13	BF686388.1
10	MARK4_H_SEQID#NA_13	BM999847.1
1	DCAMKL2_H_SEQID#NA_14	BM467980.1
2	DCAMKL2_H_SEQID#NA_14	BI034992.1
3	DCAMKL2_H_SEQID#NA_14	BI035543.1
4	DCAMKL2_H_SEQID#NA_14	BF943256.1
5	DCAMKL2_H_SEQID#NA_14	BF943502.1
6	DCAMKL2_H_SEQID#NA_14	BF362270.1
7	DCAMKL2_H_SEQID#NA_14	BF963919.1
8	DCAMKL2_H_SEQID#NA_14	BF362283.1
9	DCAMKL2_H_SEQID#NA_14	BQ217828.1
10	DCAMKL2_H_SEQID#NA_14	BF886988.1
1	PIM2_H_SEQID#NA_15	BM457909.1
2	PIM2_H_SEQID#NA_15	BM459453.1
3	PIM2_H_SEQID#NA_15	BM464831.1
4	PIM2_H_SEQID#NA_15	AU124437.1
5	PIM2_H_SEQID#NA_15	BI908737.1
6	PIM2_H_SEQID#NA_15	BI546781.1
7	PIM2_H_SEQID#NA_15	AU125921.1
8	PIM2_H_SEQID#NA_15	BI253854.1
9	PIM2_H_SEQID#NA_15	BG705716.1
10	PIM2_H_SEQID#NA_15	BM008442.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
1	PIM3_H_SEQID#NA_16	AL525596.1
2	PIM3_H_SEQID#NA_16	AL549520.1
3	PIM3_H_SEQID#NA_16	AL570770.1
4	PIM3_H_SEQID#NA_16	AL523928.1
5	PIM3_H_SEQID#NA_16	AL570076.1
6	PIM3_H_SEQID#NA_16	BI753308.1
7	PIM3_H_SEQID#NA_16	AL519345.1
8	PIM3_H_SEQID#NA_16	AL543684.1
9	PIM3_H_SEQID#NA_16	BG744856.1
10	PIM3_H_SEQID#NA_16	AL562787.1
1	TSSK4_H_SEQID#NA_17	BE551971.1
2	TSSK4_H_SEQID#NA_17	AI075923.1
3	TSSK4_H_SEQID#NA_17	BI825382.1
4	TSSK4_H_SEQID#NA_17	H87255.1
5	TSSK4_H_SEQID#NA_17	BF510751.1
6	TSSK4_H_SEQID#NA_17	BF510751.1
7	TSSK4_H_SEQID#NA_17	AW296282.1
8	TSSK4_H_SEQID#NA_17	AI218614.1
9	TSSK4_H_SEQID#NA_17	AI218614.1
10	TSSK4_H_SEQID#NA_17	AI365148.1
1	CKIL2_H_SEQID#NA_18	AL530844.1
2	CKIL2_H_SEQID#NA_18	BQ439549.1
3	CKIL2_H_SEQID#NA_18	AL577840.1
4	CKIL2_H_SEQID#NA_18	AL555305.1
5	CKIL2_H_SEQID#NA_18	AL705762.1
6	CKIL2_H_SEQID#NA_18	BE548084.1
7	CKIL2_H_SEQID#NA_18	BF433088.1
8	CKIL2_H_SEQID#NA_18	BE222107.1
9	CKIL2_H_SEQID#NA_18	AW294686.1
10	CKIL2_H_SEQID#NA_18	BG718751.1
1	PCTAIRE3_H_SEQID#NA_19	AL520700.1
2	PCTAIRE3_H_SEQID#NA_19	AL528335.1
3	PCTAIRE3_H_SEQID#NA_19	AL520699.1
4	PCTAIRE3_H_SEQID#NA_19	BM457869.1
5	PCTAIRE3_H_SEQID#NA_19	BQ437828.1
6	PCTAIRE3_H_SEQID#NA_19	BM549437.1
7	PCTAIRE3_H_SEQID#NA_19	BM045832.1
8	PCTAIRE3_H_SEQID#NA_19	BG912679.1
9	PCTAIRE3_H_SEQID#NA_19	BE747807.1
10	PCTAIRE3_H_SEQID#NA_19	BF345421.1
1	PFTAIRE2_H_SEQID#NA_20	BI755983.1
2	PFTAIRE2_H_SEQID#NA_20	BE562611.1
3	PFTAIRE2_H_SEQID#NA_20	BG326162.1
4	PFTAIRE2_H_SEQID#NA_20	AA436054.1
5	PFTAIRE2_H_SEQID#NA_20	AA435956.1
6	PFTAIRE2_H_SEQID#NA_20	BG249066.1
7	PFTAIRE2_H_SEQID#NA_20	BG249066.1
8	PFTAIRE2_H_SEQID#NA_20	BG772738.1
9	PFTAIRE2_H_SEQID#NA_20	BG720115.1
10	PFTAIRE2_H_SEQID#NA_20	W03371.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
1	ERK7_H_SEQID#NA_21	AL537138.1
2	ERK7_H_SEQID#NA_21	AL537137.1
3	ERK7_H_SEQID#NA_21	BM553342.1
4	ERK7_H_SEQID#NA_21	BM553342.1
5	ERK7_H_SEQID#NA_21	BE464560.1
6	ERK7_H_SEQID#NA_21	AI049667.1
7	ERK7_H_SEQID#NA_21	AJ403115.1
8	ERK7_H_SEQID#NA_21	AI476756.1
9	ERK7_H_SEQID#NA_21	AI921266.1
10	ERK7_H_SEQID#NA_21	AI680380.1
1	CKIIa-rs_H_SEQID#NA_22	AL559846.1
2	CKIIa-rs_H_SEQID#NA_22	AL560958.1
3	CKIIa-rs_H_SEQID#NA_22	AU131772.1
4	CKIIa-rs_H_SEQID#NA_22	AU120646.1
5	CKIIa-rs_H_SEQID#NA_22	BI258630.1
6	CKIIa-rs_H_SEQID#NA_22	AU133318.1
7	CKIIa-rs_H_SEQID#NA_22	AU133037.1
8	CKIIa-rs_H_SEQID#NA_22	AU125134.1
9	CKIIa-rs_H_SEQID#NA_22	AL582368.1
10	CKIIa-rs_H_SEQID#NA_22	AU117006.1
1	DYRK4_H_SEQID#NA_23	AL561586.1
2	DYRK4_H_SEQID#NA_23	AL582755.1
3	DYRK4_H_SEQID#NA_23	BG721331.1
4	DYRK4_H_SEQID#NA_23	BM041899.1
5	DYRK4_H_SEQID#NA_23	BI559381.1
6	DYRK4_H_SEQID#NA_23	BM042712.1
7	DYRK4_H_SEQID#NA_23	BI459242.1
8	DYRK4_H_SEQID#NA_23	BI459242.1
9	DYRK4_H_SEQID#NA_23	BF431376.1
10	DYRK4_H_SEQID#NA_23	AI066522.1
1	HIPK1_H_SEQID#NA_24	BQ224060.1
2	HIPK1_H_SEQID#NA_24	BM476759.1
3	HIPK1_H_SEQID#NA_24	BM724085.1
4	HIPK1_H_SEQID#NA_24	BG742609.1
5	HIPK1_H_SEQID#NA_24	BG681186.1
6	HIPK1_H_SEQID#NA_24	BG676057.1
7	HIPK1_H_SEQID#NA_24	AW166113.1
8	HIPK1_H_SEQID#NA_24	BG498068.1
9	HIPK1_H_SEQID#NA_24	BG612475.1
10	HIPK1_H_SEQID#NA_24	BE877361.1
1	HIPK4_H_SEQID#NA_25	BM554291.1
2	HIPK4_H_SEQID#NA_25	BG772881.1
3	HIPK4_H_SEQID#NA_25	BI827147.1
4	HIPK4_H_SEQID#NA_25	BI561789.1
5	HIPK4_H_SEQID#NA_25	BG105231.1
6	HIPK4_H_SEQID#NA_25	BG771831.1
7	HIPK4_H_SEQID#NA_25	BG720082.1
8	HIPK4_H_SEQID#NA_25	AI806773.1
9	HIPK4_H_SEQID#NA_25	AI001807.1
10	HIPK4_H_SEQID#NA_25	M62294.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
1	BIKE_H_SEQID#NA_26	BI755383.1
2	BIKE_H_SEQID#NA_26	AW968082.1
3	BIKE_H_SEQID#NA_26	BG776990.1
4	BIKE_H_SEQID#NA_26	BG485573.1
5	BIKE_H_SEQID#NA_26	AW968084.1
6	BIKE_H_SEQID#NA_26	AI939552.1
7	BIKE_H_SEQID#NA_26	AL546234.1
8	BIKE_H_SEQID#NA_26	AW967339.1
9	BIKE_H_SEQID#NA_26	BI461241.1
10	BIKE_H_SEQID#NA_26	BI461241.1
1	NEK10_H_SEQID#NA_27	AI652681.1
2	NEK10_H_SEQID#NA_27	BM976126.1
3	NEK10_H_SEQID#NA_27	AI962584.1
4	NEK10_H_SEQID#NA_27	AA954906.1
5	NEK10_H_SEQID#NA_27	BG717420.1
6	NEK10_H_SEQID#NA_27	AA889152.1
7	NEK10_H_SEQID#NA_27	AA429606.1
8	NEK10_H_SEQID#NA_27	BM976173.1
9	NEK10_H_SEQID#NA_27	AA430250.1
10	NEK10_H_SEQID#NA_27	BI462787.1
1	pNEK5_H_SEQID#NA_28	AA398536.1
2	pNEK5_H_SEQID#NA_28	AA393108.1
3	pNEK5_H_SEQID#NA_28	AI627290.1
1	NEK1_H_SEQID#NA_29	AV700007.1
2	NEK1_H_SEQID#NA_29	AV700747.1
3	NEK1_H_SEQID#NA_29	AI936517.1
4	NEK1_H_SEQID#NA_29	AV699533.1
5	NEK1_H_SEQID#NA_29	BG290898.1
6	NEK1_H_SEQID#NA_29	AV700291.1
7	NEK1_H_SEQID#NA_29	AI816275.1
8	NEK1_H_SEQID#NA_29	AV699817.1
9	NEK1_H_SEQID#NA_29	BG706222.1
10	NEK1_H_SEQID#NA_29	AW976435.1
1	NEK3_H_SEQID#NA_30	BQ432111.1
2	NEK3_H_SEQID#NA_30	BI093553.1
3	NEK3_H_SEQID#NA_30	AI971454.1
4	NEK3_H_SEQID#NA_30	AI191920.1
5	NEK3_H_SEQID#NA_30	AI659549.1
6	NEK3_H_SEQID#NA_30	BI754945.1
7	NEK3_H_SEQID#NA_30	AW043698.1
8	NEK3_H_SEQID#NA_30	AI627473.1
9	NEK3_H_SEQID#NA_30	BM984985.1
10	NEK3_H_SEQID#NA_30	AA873814.1
1	SGK069_H_SEQID#NA_31	None
1	SGK110_H_SEQID#NA_32	None
1	NRBP2_H_SEQID#NA_33	AL564934.1
2	NRBP2_H_SEQID#NA_33	BG108500.1
3	NRBP2_H_SEQID#NA_33	BQ014431.1
4	NRBP2_H_SEQID#NA_33	BQ182709.1
5	NRBP2_H_SEQID#NA_33	BG913260.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
6	NRBP2_H_SEQID#NA_33	AW962453.1
7	NRBP2_H_SEQID#NA_33	BM709377.1
8	NRBP2_H_SEQID#NA_33	BF944679.1
9	NRBP2_H_SEQID#NA_33	BG571713.1
10	NRBP2_H_SEQID#NA_33	BG576689.1
1	CNK_H_SEQID#NA_34	BG675045.1
2	CNK_H_SEQID#NA_34	BM927202.1
3	CNK_H_SEQID#NA_34	BE250216.1
4	CNK_H_SEQID#NA_34	BQ065567.1
5	CNK_H_SEQID#NA_34	BE515113.1
6	CNK_H_SEQID#NA_34	BE783099.1
7	CNK_H_SEQID#NA_34	BQ228988.1
8	CNK_H_SEQID#NA_34	BF205939.1
9	CNK_H_SEQID#NA_34	AI951666.1
10	CNK_H_SEQID#NA_34	BQ066297.1
1	SCYL2_H_SEQID#NA_35	BM905696.1
2	SCYL2_H_SEQID#NA_35	AL563032.1
3	SCYL2_H_SEQID#NA_35	AL700123.1
4	SCYL2_H_SEQID#NA_35	AU130771.1
5	SCYL2_H_SEQID#NA_35	AL528010.1
6	SCYL2_H_SEQID#NA_35	AU120073.1
7	SCYL2_H_SEQID#NA_35	BE614405.1
8	SCYL2_H_SEQID#NA_35	BM459956.1
9	SCYL2_H_SEQID#NA_35	BM786779.1
10	SCYL2_H_SEQID#NA_35	BF982530.1
1	SRPK2_H_SEQID#NA_36	BM464185.1
2	SRPK2_H_SEQID#NA_36	BQ428104.1
3	SRPK2_H_SEQID#NA_36	AL521820.1
4	SRPK2_H_SEQID#NA_36	AL045362.1
5	SRPK2_H_SEQID#NA_36	AL521821.1
6	SRPK2_H_SEQID#NA_36	AU124932.1
7	SRPK2_H_SEQID#NA_36	BG200431.1
8	SRPK2_H_SEQID#NA_36	BG389934.1
9	SRPK2_H_SEQID#NA_36	BM979654.1
10	SRPK2_H_SEQID#NA_36	AI038250.1
1	TLK1_H_SEQID#NA_37	BM561353.1
2	TLK1_H_SEQID#NA_37	AL526362.1
3	TLK1_H_SEQID#NA_37	BI488932.1
4	TLK1_H_SEQID#NA_37	AU124094.1
5	TLK1_H_SEQID#NA_37	AU119119.1
6	TLK1_H_SEQID#NA_37	BM470340.1
7	TLK1_H_SEQID#NA_37	AU134085.1
8	TLK1_H_SEQID#NA_37	BG779394.1
9	TLK1_H_SEQID#NA_37	BM981774.1
10	TLK1_H_SEQID#NA_37	BM724955.1
1	SGK071_H_SEQID#NA_38	BI458908.1
2	SGK071_H_SEQID#NA_38	AL044935.1
3	SGK071_H_SEQID#NA_38	BQ184985.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
4	SGK071_H_SEQID#NA_38	AV763896.1
1	SK516_H_SEQID#NA_39	BM918041.1
2	SK516_H_SEQID#NA_39	BQ215537.1
3	SK516_H_SEQID#NA_39	BG479508.1
4	SK516_H_SEQID#NA_39	BG475168.1
5	SK516_H_SEQID#NA_39	BG473913.1
6	SK516_H_SEQID#NA_39	BG280828.1
7	SK516_H_SEQID#NA_39	BG763842.1
8	SK516_H_SEQID#NA_39	BQ279002.1
9	SK516_H_SEQID#NA_39	BI090323.1
10	SK516_H_SEQID#NA_39	BG033116.1
1	H85389_H_SEQID#NA_40	AW292935.1
2	H85389_H_SEQID#NA_40	BG875928.1
3	H85389_H_SEQID#NA_40	BG766602.1
4	H85389_H_SEQID#NA_40	BG766602.1
5	H85389_H_SEQID#NA_40	AW027321.1
6	H85389_H_SEQID#NA_40	AW027332.1
7	H85389_H_SEQID#NA_40	AL732079.1
8	H85389_H_SEQID#NA_40	BG875945.1
9	H85389_H_SEQID#NA_40	BG875933.1
1	Wee1b_H_SEQID#NA_41	BM790836.1
2	Wee1b_H_SEQID#NA_41	BG402079.1
1	Wnk2_H_SEQID#NA_42	BQ222235.1
2	Wnk2_H_SEQID#NA_42	AL534358.1
3	Wnk2_H_SEQID#NA_42	BM907282.1
4	Wnk2_H_SEQID#NA_42	BI546992.1
5	Wnk2_H_SEQID#NA_42	BI756222.1
6	Wnk2_H_SEQID#NA_42	BM678640.1
7	Wnk2_H_SEQID#NA_42	AW962621.1
8	Wnk2_H_SEQID#NA_42	BM689160.1
9	Wnk2_H_SEQID#NA_42	BF336877.1
10	Wnk2_H_SEQID#NA_42	AI637586.1
1	MAP3K1_H_SEQID#NA_43	AU132367.1
2	MAP3K1_H_SEQID#NA_43	AL133917.1
3	MAP3K1_H_SEQID#NA_43	BM928438.1
4	MAP3K1_H_SEQID#NA_43	BM928438.1
5	MAP3K1_H_SEQID#NA_43	AL042445.1
6	MAP3K1_H_SEQID#NA_43	AW499603.1
7	MAP3K1_H_SEQID#NA_43	BG119132.1
8	MAP3K1_H_SEQID#NA_43	BE162514.1
9	MAP3K1_H_SEQID#NA_43	BF216567.1
10	MAP3K1_H_SEQID#NA_43	BF216567.1
1	MAP3K8_H_SEQID#NA_44	BM969829.1
2	MAP3K8_H_SEQID#NA_44	BG484791.1
3	MAP3K8_H_SEQID#NA_44	BI832332.1
4	MAP3K8_H_SEQID#NA_44	N57475.1
5	MAP3K8_H_SEQID#NA_44	AI683447.1
6	MAP3K8_H_SEQID#NA_44	N47620.1
1	STLK6-rs_H_SEQID#NA_46	AL552387.1
2	STLK6-rs_H_SEQID#NA_46	AL515422.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
3	STLK6-rs_H_SEQID#NA_46	AL520217.1
4	STLK6-rs_H_SEQID#NA_46	AL520216.1
5	STLK6-rs_H_SEQID#NA_46	BM465416.1
6	STLK6-rs_H_SEQID#NA_46	BI825875.1
7	STLK6-rs_H_SEQID#NA_46	BI859101.1
8	STLK6-rs_H_SEQID#NA_46	AL558687.1
9	STLK6-rs_H_SEQID#NA_46	BI823806.1
10	STLK6-rs_H_SEQID#NA_46	BI765467.1
1	MAP2K2_H_SEQID#NA_47	BQ278839.1
2	MAP2K2_H_SEQID#NA_47	AL525264.1
3	MAP2K2_H_SEQID#NA_47	BM804931.1
4	MAP2K2_H_SEQID#NA_47	BM920109.1
5	MAP2K2_H_SEQID#NA_47	BI826639.1
6	MAP2K2_H_SEQID#NA_47	AL525439.1
7	MAP2K2_H_SEQID#NA_47	BG769213.1
8	MAP2K2_H_SEQID#NA_47	BE732844.1
9	MAP2K2_H_SEQID#NA_47	BG770148.1
10	MAP2K2_H_SEQID#NA_47	BG769998.1
1	CCK4_H_SEQID#NA_48	AL515621.1
2	CCK4_H_SEQID#NA_48	BM554494.1
3	CCK4_H_SEQID#NA_48	AL515620.1
4	CCK4_H_SEQID#NA_48	BM048660.1
5	CCK4_H_SEQID#NA_48	BM801688.1
6	CCK4_H_SEQID#NA_48	AL558185.1
7	CCK4_H_SEQID#NA_48	BI871758.1
8	CCK4_H_SEQID#NA_48	BM802337.1
9	CCK4_H_SEQID#NA_48	BG773310.1
10	CCK4_H_SEQID#NA_48	BF981652.1
1	LMR1_H_SEQID#NA_49	BI603257.1
2	LMR1_H_SEQID#NA_49	BF306070.1
3	LMR1_H_SEQID#NA_49	BM549649.1
4	LMR1_H_SEQID#NA_49	BG827921.1
5	LMR1_H_SEQID#NA_49	BI601257.1
6	LMR1_H_SEQID#NA_49	BM547656.1
7	LMR1_H_SEQID#NA_49	BG911396.1
8	LMR1_H_SEQID#NA_49	AL120105.1
9	LMR1_H_SEQID#NA_49	AV727368.1
10	LMR1_H_SEQID#NA_49	BI600711.1
1	RYK_H_SEQID#NA_50	BQ067310.1
2	RYK_H_SEQID#NA_50	BQ434679.1
3	RYK_H_SEQID#NA_50	BG740085.1
4	RYK_H_SEQID#NA_50	BG764027.1
5	RYK_H_SEQID#NA_50	AU130782.1
6	RYK_H_SEQID#NA_50	BI870859.1
7	RYK_H_SEQID#NA_50	BM450529.1
8	RYK_H_SEQID#NA_50	BG762507.1
9	RYK_H_SEQID#NA_50	AL038696.1
10	RYK_H_SEQID#NA_50	BG260940.1
1	LRRK2_H_SEQID#NA_51	BG189993.1
2	LRRK2_H_SEQID#NA_51	BM998398.1

274

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
3	LRRK2_H_SEQID#NA_51	AV705213.1
4	LRRK2_H_SEQID#NA_51	BF665089.1
5	LRRK2_H_SEQID#NA_51	AW958959.1
6	LRRK2_H_SEQID#NA_51	BF699250.1
7	LRRK2_H_SEQID#NA_51	BF669643.1
8	LRRK2_H_SEQID#NA_51	BF669643.1
9	LRRK2_H_SEQID#NA_51	BQ437477.1
10	LRRK2_H_SEQID#NA_51	BQ437477.1
1	pMLK4_H_SEQID#NA_52	BG824246.1
2	pMLK4_H_SEQID#NA_52	BI964128.1
3	pMLK4_H_SEQID#NA_52	BG540713.1
4	pMLK4_H_SEQID#NA_52	BI964177.1
5	pMLK4_H_SEQID#NA_52	BE867187.1
6	pMLK4_H_SEQID#NA_52	AW408639.1
7	pMLK4_H_SEQID#NA_52	BI963837.1
8	pMLK4_H_SEQID#NA_52	BF352800.1
9	pMLK4_H_SEQID#NA_52	BI963872.1
10	pMLK4_H_SEQID#NA_52	H67242.1
1	KSR_H_SEQID#NA_53	BI086433.1
2	KSR_H_SEQID#NA_53	BF528425.1
3	KSR_H_SEQID#NA_53	AI123553.1
4	KSR_H_SEQID#NA_53	BM989782.1
5	KSR_H_SEQID#NA_53	BI091489.1
6	KSR_H_SEQID#NA_53	AW963516.1
7	KSR_H_SEQID#NA_53	AI809969.1
8	KSR_H_SEQID#NA_53	AI458861.1
9	KSR_H_SEQID#NA_53	AI088028.1
10	KSR_H_SEQID#NA_53	AW166454.1
1	KSR2_H_SEQID#NA_54	BF948353.1
1	KIAA1646_H_SEQID#NA_55	BM479389.1
2	KIAA1646_H_SEQID#NA_55	BG754980.1
3	KIAA1646_H_SEQID#NA_55	BM453176.1
4	KIAA1646_H_SEQID#NA_55	BQ230294.1
5	KIAA1646_H_SEQID#NA_55	BQ054406.1
6	KIAA1646_H_SEQID#NA_55	BQ063738.1
7	KIAA1646_H_SEQID#NA_55	BG284450.1
8	KIAA1646_H_SEQID#NA_55	BQ057191.1
9	KIAA1646_H_SEQID#NA_55	BE898542.1
10	KIAA1646_H_SEQID#NA_55	BG750642.1
1	DGK-beta_H_SEQID#NA_56	BG912323.1
2	DGK-beta_H_SEQID#NA_56	BG201482.1
3	DGK-beta_H_SEQID#NA_56	BF949285.1
4	DGK-beta_H_SEQID#NA_56	BI032910.1
1	IP6K1_H_SEQID#NA_57	AL515350.1
2	IP6K1_H_SEQID#NA_57	AL537602.1
3	IP6K1_H_SEQID#NA_57	BM544669.1
4	IP6K1_H_SEQID#NA_57	BM546339.1
5	IP6K1_H_SEQID#NA_57	BQ232298.1
6	IP6K1_H_SEQID#NA_57	BQ053969.1
7	IP6K1_H_SEQID#NA_57	AL536739.1

TABLE 6 Cont'd
Human ESTs

Rank	Gene	Human EST
8	IP6K1_H_SEQID#NA_57	BG468729.1
9	IP6K1_H_SEQID#NA_57	BG822723.1
10	IP6K1_H_SEQID#NA_57	BQ220938.1
1	YAB1_H_SEQID#NA_58	BM925217.1
2	YAB1_H_SEQID#NA_58	BI771908.1
3	YAB1_H_SEQID#NA_58	BM475528.1
4	YAB1_H_SEQID#NA_58	BM929585.1
5	YAB1_H_SEQID#NA_58	BQ067514.1
6	YAB1_H_SEQID#NA_58	AW964156.1
7	YAB1_H_SEQID#NA_58	BI829039.1
8	YAB1_H_SEQID#NA_58	BG541994.1
9	YAB1_H_SEQID#NA_58	BI871673.1
10	YAB1_H_SEQID#NA_58	BE797060.1
1	AF052122_H_SEQID#NA_59	BI833115.1
2	AF052122_H_SEQID#NA_59	BI227273.1
3	AF052122_H_SEQID#NA_59	BI226041.1
4	AF052122_H_SEQID#NA_59	BI226088.1
5	AF052122_H_SEQID#NA_59	BE562310.1
6	AF052122_H_SEQID#NA_59	BM458499.1
7	AF052122_H_SEQID#NA_59	BG324551.1
8	AF052122_H_SEQID#NA_59	BF057611.1
9	AF052122_H_SEQID#NA_59	BG779592.1
10	AF052122_H_SEQID#NA_59	BI117258.1
1	AAF23326_H_SEQID#NA_60	BI193027.1
2	AAF23326_H_SEQID#NA_60	BQ214713.1
3	AAF23326_H_SEQID#NA_60	AI907812.1
4	AAF23326_H_SEQID#NA_60	AA478358.1
5	AAF23326_H_SEQID#NA_60	AA459637.1
6	AAF23326_H_SEQID#NA_60	AA284562.1
7	AAF23326_H_SEQID#NA_60	AA401004.1
8	AAF23326_H_SEQID#NA_60	BM128250.1
9	AAF23326_H_SEQID#NA_60	BG219074.1
10	AAF23326_H_SEQID#NA_60	W88819.1
1	SGK493_H_SEQID#NA_61	BQ049234.1
2	SGK493_H_SEQID#NA_61	BM554462.1
3	SGK493_H_SEQID#NA_61	AL527674.1
4	SGK493_H_SEQID#NA_61	AL527675.1
5	SGK493_H_SEQID#NA_61	AU133075.1
6	SGK493_H_SEQID#NA_61	BI833907.1
7	SGK493_H_SEQID#NA_61	BI819237.1
8	SGK493_H_SEQID#NA_61	BM677568.1
9	SGK493_H_SEQID#NA_61	BI756523.1
10	SGK493_H_SEQID#NA_61	BG388681.1
1	BRD2_H_SEQID#NA_62	BQ222485.1
2	BRD2_H_SEQID#NA_62	BM800104.1
3	BRD2_H_SEQID#NA_62	BQ212470.1
4	BRD2_H_SEQID#NA_62	AU141190.1
5	BRD2_H_SEQID#NA_62	BQ054271.1
6	BRD2_H_SEQID#NA_62	AU143483.1
7	BRD2_H_SEQID#NA_62	BQ072172.1

What is claimed :

1.   An isolated, enriched or purified nucleic acid molecule encoding a kinase polypeptide, wherein said nucleic acid molecule comprises a nucleotide sequence that:

(a)   encodes a polypeptide having an amino acid selected from the group consisting of those set forth in SEQ ID NO: 67 though 132;

(b)   is the complement of the nucleotide sequence of (a);

(c)   hybridizes under stringent conditions to the nucleotide molecule of (a) and encodes a kinase polypeptide;

(d) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132, except that said polypeptide lacks one or more, but not all, of an N-terminal domain, a C-terminal catalytic domain, a catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline-rich region, a spacer region and a C-terminal tail; or

(e) is the complement of the nucleotide sequence of (d).

2.   An isolated, enriched, or purified kinase polypeptide, wherein said polypeptide comprises:

(a) an amino acid sequence at least about 90% identical to a sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132; or

(b) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132, except that the polypeptide lacks one or more, but not all, of the domains selected from the group consisting of an N-terminal domain, a C-terminal catalytic domain, a catalytic domain, a C-terminal domain, a coiled-coil structure region, a proline- rich region, a spacer region-and a C-terminal tail.

3.   An antibody or antibody fragment having specific binding affinity to the kinase polypeptide of claim 2, or to a domain thereof.

4.   A hybridoma which produces the antibody of claim 3.

5.   A kit comprising an antibody which binds to a polypeptide of claim 2 and a negative control antibody.

6.   A method for identifying a substance that modulates the activity of a kinase polypeptide comprising the steps of:

(a)   contacting a kinase polypeptide substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132 with a test substance;

(b)   measuring the activity of said polypeptide; and

(c)   determining whether said substance modulates the activity of said polypeptide.

7.   A method for identifying a substance that modulates the activity of a kinase polypeptide in a cell comprising the steps of:

(a)   expressing a kinase polypeptide having a sequence substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132;

(b)   adding a test substance to said cell; and

(c)   monitoring a change in cell phenotype or the interaction between said polypeptide and a natural binding partner.

8.   A method for treating a disease or disorder by administering to a patient in need of such treatment a substance that modulates activity of the kinase polypeptide according to claim 2.

9.   A method for detection of a kinase polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein said method comprises:

(a) contacting said sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of the nucleic acid molecule of claim 1; and

(b) detecting the presence or amount of the target region:probe hybrid, as an indication of said disease or disorder.

10.   An isolated, enriched or purified nucleic acid probe consisting essentially of about 10-30 contiguous nucleotide bases of a nucleic acid sequence that encodes a polypeptide selected from the group consisting of SEQ ID NO: 67 through 132.

11.   A recombinant cell comprising the nucleic acid molecule of claim 1.

12.   A vector comprising the nucleic acid molecule of claim 1.

13.   A method for identification of a nucleic acid encoding a kinase polypeptide in a sample, wherein said method comprises:

(a) contacting said sample with the nucleic acid probe of claim 10; and

(b) isolating a nucleic acid that hybridizes to said probe, thereby identifying said nucleic acid encoding a kinase polypeptide.

14.   A transgenic mouse comprising a nucleic acid sequence that encodes a polypeptide substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO: 67 through 132; wherein said mouse exhibits a phenotype, relative to a wild-type phenotype, comprising modulation of kinase activity of said polypeptide.

15.   A cell or cell line obtained from the transgenic mouse of claim 14.

Figure 1

>CRIK_H_SEQID#NA_1

```
GTCTCGCCAGGTGGTGGAGCTGGCAGTGAAGGAGCACAAGGCTGAGAGATTCTCGCTCTGCAGCAGGCTCTCAAAGAGCAGGAAGGCCGAGAGCCTCT
CTGACAAGCTCAATGACCTGGAGAAGAAGCATGCTATGCTGAAATGAATGCCGAGAGCTTACAGCAGAGCTTGAGACTGAGACTGCAAACAGAG
GCTTCTGGAAGAGCAAGCCAAATTACAGCAGGATGACCTGCAGAAAATCAGATTTCCGTCTGACTCAAGAAGCTCTAGATCGGGCTG
ATCTACTGAAGACAGAAGAAGTGACTTGCAGCCAAATGACCAACCTGCTATCTCATTCAGTTCTCATTCATGAAAAGATGAAAATGGAAGGCACTATTTCTCAA
CAAACCAAACTCATTGATTTTCTGCAAGCCAAAATGGACCCTTCAAAAGAAAAAGGTTCCTCTGCAGTACAATGAAGGCTGGCCTGGAGAA
GGAGAAGCTCGCTGTGCAGGAGCTAGAGGAAGCCCGCACCGCCAGCGAGATCGCCATGTGCCAGAGACCACCAGCGAAAGCAAGGAC
CACCCACACACCGCCAGCCATCCAGCCGCGAGAAAGGAGTCTTCAACTCCAGGAATTTAGTCGGCGTCTTAAGGAACGCATGCACCACAATATTCCTCACCGATTCAAGTAGG
CCGCCATCCAGCCGCGAGAAAGGAGTCTTCAACTCCAGGAATTTAGTCGGCGTCTTAAGGAACGCATCCAAATGTCTCTGGATACCCGTGACTTTGGATATCCGTGACAAATGTCAGGTGATGTGTCACCCCAA
ACTGAACATGCGAGCAGCCACAAAAGTGTGTCAGGCCACCTGGCTCTGCTGTGTGGCCTTGCCACCTGGAAGGTGGCTGCCAGGAAGTACAGTGTGTCACCCTCCA
GTGCTCCACGTGCTTGCCAGCCAGCAGCCAGCTTGCACCTGGATCTGGATACCCGTGAATATGCCACCTTCACCGAGGCCTTCTGCGTGACAAAATGAACTCCCAGGTCTCCA
GACCAAGGAGGATCAAAAGTCCTCATTTATGACAACATTCCTGAACCTGCCCAGGGTGGACAAGGACCTGGACAGCAAGAAGTACATTGTC
CTGGAGGGGAAGACCTGGAAACAAAGTCCTCATTTATGACAACATTCCTGAACCTGCGGTGCCATAGCCAAAGCAGAGTGTCCCATACACTGTCCTGCCC
TATTCATGGTGCCGTTGGTGCTTCCGAACTTGCTAGCTCCCAGCTTCCCTGACAAACAGCAGCGCTGGATCCTCACCCGCCACACCTGCTGCC
CGGGAGAACCCTCTACTTGCTAGCTCCCAGCTTCCCTGACAAACAGCAGCGCTGGATCCTCACCCGCGTGGGAGAGTTTCTAGGGAAAA
AGCAGAAGCTGATGCTAAACTGCTAAACTCGCTACGCCCTACAACCTGCTACAGTGTCTCCAAATTTATATTATCA
TGTTGGTGGGCACCGAGGAAGGGCTCTACGCCCTGAAATGTCTTGAAAACTCCCATGTCCCAGGAATTGGAGCGATGGAAAACAGTCCCTGCCCCAGTCCCACCTGCCT
AGGACCTGGAGAAGCTACTCATGATAGCAGGAAGAGAGCGGGCACTGTCAAGGGCTGCCACTTGTTGGGCAGGCAGATAGAGACCTCAGCAGTCTGATAAGAATGACCATTCCTTGGCACC
GCCCAGCCCGACATCTCACCCAACATTTTGAAGCTGTCACAACGAAACCTCAGCAAATACTGGCAAATGCAAATGACCTCAGCGTTTATCACTTCAC
CCAGCAAAGTCGTCATTCTCCGCTACAACGAAACCTCAGCAAATGAAATCGACAGTCGTGCAGTCTCAATCGGTCAGGGCTGTGTGCATGTGTTCCACGAATTGG
CAATTACAGTATCCTACAACCAATAAATTCTACAAATGAAATGCTACGAAATCGTGACAGGAGCAGTCAAGTGCCTTTGGCCTTTGCCTACAGAGAACCCTATCTGTTTGTGACC
TGCTGTGTTTGCCGGTGATTCTTACGGAAGACGTAGCCGCACAGACGATCTCAGTGAGTGCTTTGGCCTTTGCCTACTGGAGCGTACCTGGGCCCT
AGTGTTCGTGATGAATTCACTCAGGAGGCGATTAATTGACACTCAACGCCATCAACCTCGTACAAGGGTCATTCATCAAAGAAACCTCGTGAAAGGAGTCGGGCACTGAACAC
CACCGGGGCCCCAGCCCACCTGCGGCGCCAGCCAAGCCAAGAGCCAAGCACACCCACCGCGAGAGGGGACGTCCCCGGAGAGGCTGTTGAAGACGCAGCAGCAGCGCGCCCTGGAGCCG
CGAAGGCAAGTCCCCGGCCGGCGGATCGTCAGCGAGATGCTCAGCAGGAGGGGTCCCGGAGAGGCTGTTGAAGACGCAGCAGCAGCGCGCCG
CGAGAGAAGTCCCGGCTGTCCCAGGTGAACAAGGTCTGGGACCCAGTCTTCAGCCAGAAAAACCAACTCCTCATCTGATCTGCAGAAAAACAC
TGAGGACCCGGCTGTCCCAGGTGAACAAGGTCTGGGACCCAGTCTTCAGCCAGAAAAACACCAACTCCTCATCTGATCTGCAGAAAAACAC
CAAACACACTATGGAACTCTGCTGCTGATGGGAACCAAGGTCTGGGACCCCAAGGGCCCCACCTCGGCCACGGACACCCC
TGTCTCCAGGAGGGGCAGGTGGCTGAGGCTTGGTTGCTACCTAGCTCCAGTTAGTTCATCAAAGGAAACCAACCGGATGCTAACTACACATGGTTAGAATCCTA
TTCAAAGCATTCACAACTTTTGTACCTAGCTCCAGTTAGTTCATCAAAGGAAACCAACCGGATGCTAACTACACATGGTTAGAATCCTA
ATTAGCTACTTTGAAGATCCTAGGATTGGTTGTTCGTGTTTTTTCTTCGTTTTTTTTTTTTAAGACAACAGAATTCTTAATAGATT
TGAATAGCGAACGTATTTCCGTGTTAGTCATTTTTAGCTGACCATCATCAGGTCTTTGCCACGGCATAGTGTAGAAGAATTCGGTCAGTTGGCCAA
CCTCCCGGACCAAGTAGGTTCATCCTCCCACCAGTGAATGTATTGAGAAGTGATGAAGTCGGGACCTCAGGGGGACCTCGAG
CTTTTTATGGCCTTGACCTTCCCCTCCACCAGCGTCATCTCCCAGCCTCGGGGAGTCTCATAGCTTTGGCGGTGTTGTTCCCATATAGCTTTGGCGATGGGGAAGAGGCCCGTC
AGCACTTCCTTCCCCACCCACCCCAGCGTCATCTCCCAGCCTCGGGGAGGTCTCATGGCCATGTGCTCCATGTGCTCCAATGTATTAATACCATCCCATAGCTCTGCCAAATGCGACACCCTCTGAC
GGGAGTTCTGCATGGCAGTTCACTGCAGGTTGCTGCCAATGTCCTGGTTCTGCCAATGTATTAATAACGCTGAAACAAACGCTCAAGAAGGACCCCTCTGAC
GACTTGCCGACTAACTGGCACCACCAAGCTGCGACTAACTGGCACCACCAAGCTTCAGTCTGTGAACAAACAAAAACGCTCAAGCCTTACGACCAGAAGAAGGATTTCAGCAAA
```

CCACCACCTCCCACTCAGTGTCCCCTCCAAACTTCACACTTCCCTGCCTGCAGAGGATGACTCTGTTCACACCCAATCCAGCGGCGGTTCTACCCCACGAAACT
GTGACTTTCCAAATGAGCCTTTCCTAGGGCTAGACCTAAGACCAGGAAGTTTGAGAAGCAGCCGCAGTCAACTCTTCAGCTCCGCCAGGGTTGGGAA
GTCCTTAGGTGCAGTGCGGCTCCCACTGGGTCTGCGGACCCTCCTATTAGAGTACGAAATTCCTGGCAATCTGGTATAGAAACAACCTAGAGGCTTTGCAGTT
GGCAAGCTAACTCGCGGCCTTATTCTGCCTTAATCTCCCACAAGGCATCTGTTGCTTGGGTCCTCCACGACTCTTAGGCCCGCCTCAACAACCCAGGCAC
CTCCTAGGTAGGCTCAAAGGTAGACCCGTTCCACCGCAGCAGGTGAACATGACCGTGTTTCAACTGTGTTCCACATGTGGTAGCTCCTTTGGCAAATGTCCT
CTGGCCTGCCATCCCAGCTCCTTCCTGCTCGTGTGTTGCTTGGGGTCGTTGTGGGTCGTTGGGGTCGTGTGGAGTCTGGAGTCTGGGGTGCTCTGTTTGTCTCTTG
GCTGTGGCGTTTTATGTGTGCTTTTATGTGTCACTGAAAATGGGTCTTTCACTGCTCTTCTGTATAGAAGGTCACTGCTCTTCTGTATATAGAAGGTCATAGAAATGTATA
GTCCACAGTATTTGGAAAGGTCACTGAAAAGGTCACTGCTCTTTATATTGCTATATTGGGACAGCCATCTCACTGAGGGTGG
TTGTGTGTCTCATCTGTGACCAGATCTTTCTGTATAGAACTGCTTTTTATGTGTAAAAGGTTTGTTCAATGGTGGTAGTTGTGGACAGCAGTTTTCACTGAGGGTGG
ATTGAATTGGTTTCTACTTTTTATTGTCATTTTATGTGTAAAAGGTGCATTTTCAACACTTACTTTGTTCAAAGCTTATTAGCATCAATCAAGATTCTAGGTCTCCAAAG
GGAGTCCGTGTGACCACCCAGATGCCAGCAGGATATACCGTAACACGAAATTGCTGTCAAAAGCTTATTAGCATCAATCAAGATTCTAGGTCTCCAAAG
TACAGGCTTTTCTTCATTACCTTTTTATTCAGAAACGAGGAAGAACACAAGGAATGATTCAAGATCCACCTTGAGAGGAATGAACTTTGTTGTTGAACA
ATTAGTGAAATAAAGCAATGATCTAAACT

>DMPK2_H_SEQID#NA_2
CGGCCTGACAGGCGGGCATGCGGGGGCGCGAGCAGCGAGGCTCCGGCCGGCTGCGCGGCGAGCAGCTGGC
GCGGGGGCGAGGCCGGCCGGGGCTGCCGGGCTGCCGGCTCCTCCTAGATCTGCTGCTGCGGCGCTCAGCAGCTCAGCAGCGGGGAGCGGCA
GCGTGGCGCAGTTCCTGAGCTGGCCAGCCCTTCGTATCAAAGGTGAAAGAACTGCGTCTGCAGAGAGATGCTGAAGGTGATCGGCCGA
GGAGCCTTTGGGGAGGTCACCGTGGTGAGGGGGACAGCCGTTGGGTCGTGAAAGGGGACAGCCGTTGGGTCGTGAAAGGGGAGTACCTGTACCTT
CAGCCTGTTTCCGGGAGGAGCGGGGATGGTCTCGTGAAAGGGGACAGCCGTTGGGGTCTGAGGACCGGTTGGGGTCTGAGGACCTGCGGCTGGGTGACTTC
GTGATGGACTACTATGCTGGTGGGGGGACCTGGGTTAATGTCCACAGGGATGTCAAGCCAGAACAAGTTCTGTGATGTCAAGCGGCACATTGGCTGGCTGGCTGACTTC
CTGGCCCATCCACTGCGTCTCAACAACCAAGGGCCCACAGTGTGGTCACCCGAGATGTCAGTGGCTGGAGTCTTGGGGGAGTCCCGAGTCCCTTGGT
GGGCAAGGGCCACTACGGCCAAGATCATGAACCACGAGGACCACCTGCAGTTCCCCCGGAACCATCTTTCTTCTTCTTCTTGAAAGGGGTGGTGACGTGCGCGAGCAGCACG
TGTGTCGCCAGGAAGAGCGGCTAGGCGGTGGGGGGCGCCCATGACACCTCCAACCTCAACCATCCAGGAGCAGTCTGAGGCTTGGCCTGCCTGGAGCGGG
GGGGCCTTCCCGGCCATCACCTGTGGCTTCACCTGTGGCTTCACCTACACCTGCACCTCACCTCACCTCAGGGAGCTGTGAGCAGCTAGCAGCTACGGA
AAGCTCCAGTGTCTGGAGCAGGAGAAGGTGGAGCTGCCAGAGATGCTGAGGGACACCAGCAAGGCCTCATTGTCCAGGGCCATGGGCCCCAGGTCAGGA
AGGAAAGTGCAGACTACGGCAGGAGCTGCGGGGACTTCACCGGAGCTGGCCGAGGTCTCAGGGACCTGGAGCTGGAGGCTCTGCAGGGGCCCAGGGG
CAGTGACCTACGGCAGGAGCTGCTTGACCGAGCTTCACGGAGGCTACAGGAGGGCCCACACGTGGGACGCAGTGGGACAGTCGTCCCA
GGGCTGCCCAGAGGGGAGCTGGAGGGAGCTGCGGGGCCCATGACACGGAAGCAGAGCTGGCAGGTGCGCCCTGGA
GGCCAAGACCATCCACACAGCCTCTGAGACCACAGCCACAGCCCACAGGAGGCGAGCTGGAGGGTGGGCCGCCCTGCGA
GAGCAGCTGGAGCAGTAGCAGCCCACACGGCCCAGGAGCAGGAGGGCTGGAGGGTGAGGGTGGGGGAGCCTGGAGGGTGGGGGAGCCTGAGCTGGAGGCCCGCGAC
ATCCTCAGCTGGGTGAAATGAGGAGCTGCCGGCCCACTCAAGAGAAGGTCTCAAAGAGGCTACTCGCCAGTGGGGGCCTAGCTGAGTCAGGGCCTAGGCCCGCGAC
CCAGACGCCTCCCTGCGCCCAGCTGCAAGATGGAGGCCGTGAGCTGCAGCTGCAGTCGGAGGCCGAGATCGCGTGGAGGCCGAGATCGGAGGCCGAGAAGCCAGGGGGCTCGCCATGC
GAGCGGGCTGACACAGGTGCACAGAGGAGCGCGAGCAGCTGCAGGCCCTGCAACACAGGAGCCCTGCAACAGAGAGCTCGCCATGC

TGCGGGAGGAGCTGCGGGCCGAGGGCCAGTGGACACCAAGCCCTCAAATCCCTGATTCCTTCCTTCCTGTCCTTCCGGAGCTCAGAGAAGGATTCTGCCAAG
GACCCTGGCATCTGCCAACACAGCCTTCTGCTAAGGAGGAGAGCCAGATCTGAGGCCGGAGGCCAGCTGCGACGCAGCGCAGCCTGCGCAGCTGCTGTTCCCAGAG
CACCCACTGCCAACACAGCCTTCTGCTAAGGAGGGGCATGGGGCCTTGGGTAATGCTGTCCCCTCCCAGCCG
GCTCACACACGCTGCGCCCCCGGAGCTTCCCATCCCCACAAGTGTCTCGCTGCACCTGTGTGATG
CCTGCGGCTACTTTTGTCACACAACCTGTGCCCGCGGCCGCCCAGTGTCCGGGGCGGTGTCCGGGCGGTGTCCGGAGTACACCCGAAACAG
GCACAGGCACTGCCTATGAGGCTTTCTGTGGGCTTCTGTGCGGCTTCTGTGCGGCTCCGAGTGACTCACGCCTGC
TGCTGTTTGACGCCCTGACCTGAGGCTCAGGTCCTAGATCTGAGGGACCCCAGTTCTCGCTACCCTGTCTGCTGCT
CTCTGATGTTATCCATGCCCAATCCAGGGACCTGCCACGCATCTTTAGGGTGACAACCTCCCAGCTGTGCTGCTGCT
GGCAGAGAGCGGAGGGGAACGCTGGCTGCAGGTGCTGCTGAACGCTGGTGCTGCTGCACGCTGCTCTCGACACGCTCTCGACACGCTGCCATCCTGCACCGCTGCAGCCGTGTACACA
CTCAAGGAGGCTTACGACAAACGGGGCTGCTCAACCGGCTCGCTCACACGCTCTCACACGTGGGGGAGTGCCGGCGGGGAGTGCCGCCTGCTGGTCGTGCTGCTG
TTTGTCATCCATCTGCGCAGCAACGACACATCTTCCAGGTGGGGGAGTCCCTGGCCGGAGCTCGAGGCTCTGAGGGCTCTGAGGCTCTGAGGGCTGCCAGGTGCT
TGTGGGCCCGCGCCCCAGCCGTGCTCTTTGCCCTGCCGTCTCTTGTGCCTGGCGAGTCTCTGTGTAGCCCGGTGCTCTGTGTAGCCCGGTGCTCTGTGTAGCCCGGCC
GGCAGCTGGAAGCATCCTGCAGCTGAGCTGCAGGCACCTGCCACTGTGAGCCTGGGGCGTCTGGGGGCGTATGTGTGGCGGCGCCGGTGCTTTGCACTCTACC
AGCGCCGCATCCGTGAGCTGCGCGCCATCCGTGAGCTGCGCGCCATCCCGCGGGGGCCTGGGTGAGGCACTGGGTGCCGTG
CGCTGCTCAAGCGAGGCTGCGCGTTGGCGCCGGTTTGGTGCTGAGGAGCTGCCACCATCCCGCGCGCCAAGTCTCGTGGCCAGCGCCACCAGTCTCGTGGTGTGGCCAGCA
GAGCTTAGCCTCAGCGAGTTCTGCTGCTACTTCACCACTGCTGACAGTGTTCAGCGAGAACTCCATGCGATGTTTGACGTGAGGAGCTGCAGAATGGGTGCAGACCGT
GCGCCCCATGGGGTATGCGGCCCCTCAATCCAGAGGGCTCCGTTCCTCTACGGCACGGAAGTCCGCTGACCTACTTCAACCAGTTACTACCACCAGTCTGGCCCTGCC
GCGGCTCAAGAAGGTGCGGCGGGGCCCTCGCCTCAACAGCCGGACCTCACCTGTTCCGGCACCAAGAGCAAGGCCGAGTTGCCCGGCGCTCCGG
AGGACGAGTTCGACATCCCGGACCTGAGATGCTGCAGGAGATGCCAGGGACAGCCCCGGTCCCCGTCCCGCACTGTCACCCAAACAGTCTGGGCCAGCAGTGAAGA
CAGAAGCAGCAGCCCCGGCCCCGGCCCAGGAGATGCTGAGGGCAGGACAAGTCCCGTCCCGACTGTCACCCAAACAGTCTGGGCCAGCAGTGAAGA
AACGGGCGGCCCCGGCCCAGGACAAGTCCCGAGGCGTTGCGGCGCCCAGCCTCCATGGGCGCCCAGCCTCTCCGGACGCTCTCCGGAGCAGTCAGTGAAGA
CCCACAGCGGCCCCACCAGCTTCTCCGTGCCAGCGAGTCTGTTCCGCGCCCTCCAGGGATCGTCGAGCCCTGCAACCTCAGGGAGGCAAGAGCTTGAGGA
GGAAACCCTGGACATCCCTGAATTGGAGAGCTCTCCTTGAATCGGACATGGAAATTCGGACATGGAAATTCGGACATGGCCAATCCCAGCACCCGGGTCCCAGGACTTGGGG
TCCCCGTGTCCCTGAATTGGGCTGTGTCCGGGACATGGCTTCATCTCGGCTGGTGTCCGGGACATTGGGACACTTGGGCAATGGACTTGGCCTTTGTTCTGGTCTCCAGGG
ATGCCATACTCCGGCTGTGTCCGGGACATCTCGGCTGGTGTCCGGGACATTGGGCAATGGACTTGGGGAATAGAATGGCCCCTTTGTTCTGGTCTCCAGGG
GTCCTGACTCCCAGCCCTCATCTGCCTTACCCCTCTGTTCCACAGCCCCAGAGGCTTGTAAGAGTTGGGATGGAGGGAGGATGGAGGCTCTGGGAACAGGTTGG
GTGATTGTGCCAAAGCTCTTATTTCCAGTGCCAAGCCTGAGCCCAGTGACCCCGTCTCTGGGTGAGAAAAACAGGCTGTGATCCAGA
AGGTCTTATCTGTGGAGGAAGATAGAAGGAAGGAAGGATGTAGGCGAAGGAAGGATGTAGGCGAAGGAAGGATGTGGTAGGAATTGGAA
AGAAGGGAAGATAGAGAAGGAAGGAGGGAGGAGGAAGGGAGGGAGCGATGAAGAAGGAGGAGACAAAAAAGGGAAG
GGAGGGTAGAAGCCGTGCAGAGGGAGGGAGCTGCAGAGGGGAG

>MAST3_H_SEQID#NA_3
GTACCCGCCGCACAGAGGCGGCCTCTGCCTCGGCATGAAGTCCCGCAGGGACAAGTCGCACATCCCGGCGCTGACCCTGCGACCCTGATCTGTCTCCGAGCAGCCAG
AGCCCGTCCGGTGCTGGGTCCCAGCAGCCCCTGTAGCCCCTGCAGCCCCTGGCGCTGCACCCTGGAGCTGCGCCAGCGGGAACCGGCCCAAGAGAGCTTGGTG
GTAGGAACGCCCCTCCCCGACCCCTCTCCGCGGAGGGCAGCAGGGCAGAAGATGGTCCCTGCTATGGAACCAACACCAACAGCTCCACCC
CTGCCCTAAATTTCCCCTTTGCCCGGAGGGCAACGTCTCCGGACGCTGCCGACGAGCTGACCAGCTTCCGTCCCAGCGAGCTCAGAGA
TCTGTCAAGCTCATCTCCCGGGAAGGCGGCGGTCCCGTCACCCGCCGCTCCGACCCGCTCCGGCCGTGCAAGCGGGGACCTTGCAACTGCAGCGGTGATCCAGA
ATGTGCTTGATGAGGAAGGCGGCGGCCGTCTCGCAGTGGCCGGCCTACGGCGGCGCCACAGCGGGGCCTACGGGGCCCGGGCTG
TGATGAATCACGTTACCGGGAGAGGTTCCCGACAGATGGGGCCGTCTGCGAGGAGGGCGGCGCTGGAGGAGTTCCTGCGAGGAGTTCCTGCGAGGGGGCCGGCTG

```
TTTGCAGGCCCCTGCCCTCTGGGCTGAGAAGGATGCACTTTGGACAAGTCATCTGTGTTTGTGTTTCGTACTTTTTAAGTGTTTTGTGTTACCT
GGTCTCATTCCCCTCCCCACACCTACCCATTTGAGGGGATGGAGTTGAAGTCACCTGGTCACCTGTACCGGCCAGTCTCACCGGAGTGTCCGTA
AACAATTCCTCTCACCCACAAAACAATGTAATCCCAGCGATGGACTGGATTCTGAAGGCCACTTCCACCATCATAGCTGCCATGCCCAGTGCCTGCT
CTATATAGAGTCTGCCTCCAATCCTGCTCAGCTGCTGCGGCACTGATAAAGGAAGTACAGCCTCACGTGAACTGCCAAGGCAGCAGCTCCAGGGGC
GAGGGGTCATCGGCTCCCAGTGGCACTGCGCGGATTACGTCCAGCTACTTCCAGAAACACTCAGTGTCCCCTCCCAGCCTCTGCCTTGTCCAGTCT
ACGCTCCAGTGTATGCCAGTCCTGCGGGATTTGTTGGAGGCCCAGCAGAGTAAGCCCTTGGCCGTGATGTGTCTGAAACACCTGTTAGGGGTTCCCTCCATATGTCAGAGCCTT
ACCCTGGGGATGAAGATGCCGTGTGTTGAGGCCCAGGCTCAAGGTCAAGAAGTTCAAGGCCAGAAAACCCAGTCAATCCCATTTTACACCTGTTTATTAAGCACCTACTGTTTATTAAGCACCTACTGT
CTCTGGGGATGAAGTTCAAGGCCAGAAAACCCAGTCAATCCCATTTTACACCTGTTTATTAAGCACCTACTGTTTATTAAGCAGCTAC
AGTTTCCTTGTTTGTGTCATCACCTCCATTTTACAGACAAAGCTGAGAACCCAGGAGGGGCGGGAGACCTGTCCAGCATCACCCAGCAGGAGTCGTGGCA
GCTGGAGGTGGTGTCATCAGCCAGCGATCAGCCAGACCCTGTGTTGTGGCGTTGTCATCGAAGGGGCGTTCATCAAGGGAGGTTCTGGTTCATCAAGGGAGGTTCTGTGTGTGTG
GAACGGAGCATCAGCCAGACCCTGTTGTGGCCACTCGTGTGTGGCCACTCTGTGACCGAGCATCAACACAGCCCACTCTTCCTTTCTGAGTCCCCTGTCCAGCACTGCCT
CCCATGATGCTGTGTGACCCAGCTGGGCACTGTGCATGATTCACCCCCCTCCATCCTGGAGCTGGCAGTGAATAAAAGCCCGTATTTAC
```

>MAST205_H_SEQID#NA_4

```
ATGAAGCGGAGCCGCTGCCGCCGACCGAGCCGCCGGGGAGGATGGAGTTCAGCGGCAGCGGCAGCGGGGAGGAGATGAAGCGGAGCCGCTGCCGCC
GCGCCGGGCCGAGGGCCGAGGAGGAGCTGACGGGGCCCCGCGGGCCGAGGAGGAGGGCCGAGGAGGAGGATGTAGTAACTGGAGTTAGTCCC
CTGCTCTTCAGGAAACTCAGTAATCCTGAGCAGTATTTCATCCACTGGAAAAGTTAAACTTCAGGATGATTGTAAGTTATGGAGAAGCATCA
AACCTGGCCAGTCTCTATCGGGTAAGCAGTCGTCCCTTGTCCACAGTGTCCTTCTTCTTACTGCGCCTGGACAGTGGGACAAGTAACCGCA
AACCTGGTTCGAATGAGAAACCAGTCCCTTGGACAGTCCCTTGGACACTCCACCACCACCACCACGGCCACATCACCACCGCCACACAGGTAACCAGTTCTTTGGGACAGCGCCCCGGAATT
AGAGCTGATTGTGACCTTGACTCTAGCACATCACCTAGCACACTCTAGCACACTTTTCTTTGTTCCTGACCTGAAAAGCTGCCTTTCCAGCTGCCTTTCCAGCTGCCTTTCTTTGACGAAGCATT
TCTCTCCAAATGCACCTGTCTCATCATCATGTGGAGCAGGAAAAGCTGCCTTGCCTTCCTTCCTTTCCAGGATTTTTGACGAAGCATT
ACTCCTAGCTCCACTGTCTCATCATCATGTCCTGCACAGGCAGGAAGCGCCACACGTCAGTGCCTTCCAGGAATATGAACTAAC
TCAGCACAGAGCGTACCAGATGAGGAAGCGTACCAGATGAGGAAGCGGCAGTCCCAGCCTCAGTCGTCCCGGACGATCCCCAGTATCCTTTGAC
AGTGAAATAATAATGATGAATCATGTTTACAAGAAGATTCCCAAAGGCCACACTGGTCAAGAGCACTAGCCAGAGTTTATTTCTCCAACACTCC
AGACAGGTGCTGCCCTTGGCAGATGGAGCCCTTGCAGATCTCAAGAGAATTTGGAGAAACTTTACAAGTTGGCTTTTGTGATGCAGCT
TACATCACAATACTTCTACGAACTTCAGGACGCTCTCCCGCACTGTCCCCGCTACATCGTTCCCCGCTACATCGTTAGAAGCAGTGGAGG
GGTGAAAAAGCTGATGATTATCATTGCCCGCCAGCACGTCTCCCGCTACATCTTACCCCGGATTAAATGTGACACCCGGATTAAATGTGACAGCCCAGTTGA
CCACGCCAAAGAGGGACAAGAGGGACAGAGCTCCAGACAGATGATTCATTGGGGCATGTGAGGGCCATGGGCCATGGGCCTTGAAGAGGACTTCGAG
GCAGCTGTGACAGTCGATCATCAGCAATGGCCTATGGGCGCCTATGGGCGCCTGTATTTCTGTGGGCGCCTGTATTTCTGTGGGCGCACAAGTCCACCCGGCACAAGTCCACCCGGCACAAGATCAACAAGCAGAA
ACCATTAAGCTCATCAGCAATGGCCCTATGGGCGCCTATGGGCGCCTGTATTTCTGTGGGCGGCACATGGCATTTCGTCCTTTGATACC
CCTGATCCTACGGAACCAGATCCAGCAGGCCTTCGTGAGCGTGACACTGACTTTCGCCACTCTGCTGAAGGGACACTACTGTGCGTCTCTATAC
AAGCGCCACTTGTGCATGGTGCTGCGCCCTGGACTACTTACACAACTACATCCATGGTGGCGTCACATCATGGTGGCGTCACATC
TTTGCGGAAACTGTGCTGCTGGACTTTGGACTTTGTCCAAAATTGGCCTCAAAATGGCCTCCAAAATACATTGGCCTGCCAAAATTGGCCTGGACAA
AAGCTCACGGGACTTTGGACTGTCCAAAATACATTGGCCTCCAAAATACATTGGCCTGGGAATTCCTGGACAATCCTGTATG
GCAGGTATGCGGGGACCCCAGAATACATTGCCTCCAAAATACATTGGGGCCATGGGCCTGTGGGCCTATGAGGTTGGGGCCTGAGGGTGATGAGGCAC
AGTTCCTGCGTGCGTGCGTCCCTTTTTTTGAGATACTCCGGAGGAGCTCTTTGGAGATACTCCGGGAGGAGCCTATGAGGGTGATGAGGGTGATGAGGCAC
TGCCCCCGAGACGCGCCCAGGACCCCAAAATGTCTCCAAAACTGCTCCAGAAACCCTCTGGAGGAGCTGAATTTATTCCTCAGTTGGAGTCAGTGAGGCAGCACCCATTC
TTTACTGGTCTGGACTGGACAGGACTCTCCGCCAGAAGGCTGAATTTATTCCTCAGTTGGAGTCAGAGGATGATACTAGCTATTTTGACACCCGGCTCAGAG
```

CGATACCACCACCATGACTCGGGAGGATGAGGAAGAAGTGAGTGAGGATGGCTGCCTTGAGATCCGCCCAGTTCTCTTCCTGCTCTCCAAGGTTCAACAAGGT
GTACAGCAGCATGGAGGCGGCTCTCACTGCTCGAGGAGCCGGGACCACCACCCCGAGATATTACGGAAGCGGCTGTTGCTGAGTCATCCACACAGAGAGTGA
CTGGCAGGGCCCTCCAAATGACAGTGCGACGCGGCTCAATGCTGTCTCAGGCTGGGTGATTGGCTCCCTGGATGCGGCCTCGGTTCCCGGAGGGCCTGAGGAGGC
CTCAAGCACCCTCCAAATGACAGTGATATGGGTCCTGACACCCCGTCCAGTCCAGTGGTTCCAGTTCCAGTCCAGTTCACTGAACACTGCCAAAGCTGGATGA
AACCACAGGAGGGTATATGGGTCCGGAGCAGTGGTGTTCCAGTTCCAGTCCAGTCTATGGAGACCGAGGCCGTGGACCTCACAGCTGGCTGACAGGCCATCAGTG
GGAAGCTGTTGGCCGGAGCAGTGGGCCCGCCACCGGTGGTTCCACCGGCCACCGGTGGTTCAACAGAGAAGCGCACTGCTGCCCCGTCAACAAAGTGATCAAGTCCGCCTCAGCC
ACCTGGCTGTGCGTAGGGCCCGCCACCGGTGGTTCCACCGGCCACCGGTGGTTCAACAGAGAAGCGCACTGCTGCCCCGTCAACAAAGTGATCAAGTCCGCCTCAGCC
ACAGCCCCTCCAAGCAGGGGACTTCTTGCCAGATCTCTCGGGGGACTCAACAGAGAAGCGCCAGCCCATCATCCTGCAAGAAGTATGGCTTCACCCTGCAAGGTGAC
CTTCTCCAAGCAGGGGACTTCTTGCCAGATCTCTCGGGGGACTCAACAGAGAAGCGCCAGCCCATCATCCTGCAAGAAGTATGGCTTCACCCTGCAAGGTGAC
GCGTCTACATGGGTGACTCGGATGTCTACACCTGTCACCATATGGTGTGCACGTGGAGGATGGAGCTGATCCTGAAGAGTGGAAACAAGGTGGCCATTCAACAAC
CTCATCACCCATGTCAATGGGAACCTGTCACACATGTCACCATATGGTGTGCACGTGGAGGATGGAGCTGATCCTGAAGAGTGGAAACAAGGTGGCCATTCAACAAC
TCCCTGGAGAACACATCCATTAAAGTGGGGCCAGCTCCTGTTCCGCAAGATCCACCAAGCAAGCAAGCATCCCTGTCCACACGAGCGGCCTTTCCTTCCTTCCTTAACGCTCCTGTCA
CAAGAAAGCAGAAAAAGGAGCTCCGTTCCGCAAGATCCACCAAGCAAGCAAGCATCCCTGTCCACACGAGCGGCCTTTCCTTCCTTCCTTAACGCTGTGTGCATTCAGTG
TCAGGGGAGAGTGGGCCAGGCTCTCCACACAGACCAGCTCCCCAGTTCCCCAGTCCCCAGTTCCCCAGTCTCTTCCCCAACCCCCAACACGGCTCTGGCACCC
GGAGGGAATTCATCACAGAGACAGCTCCGCTCTCCAAGCGGCGCAGTGTCAGCAGTCAGCCAGGCCTTCCACAAAGCTTCACTTCCACAAAGCTTCCAACAGCTTCA
AAGCTCCAACGCAGTACCGCTCCCCATGTCCCCATGCTGGCCATGTAGCCGCAGTGCAGTGAGAAACTGCCGCTGAGAAACTGCCGCCCTCGGAAGAAGCTAGCCACTTCTCGC
CCTCAGCGGTCCCCATCCCCGTTCACCACTACTCACCGGTGCAGTGCGCTGAGTCGGCTGAGAAACTGCCGCCCTCGGAAGAAGCTAGCCACTTCTCGC
GTGCGGAGCCACCCCGTTCACCACTACTCACCGGTGCAGTGCGCTGAGTCGGCTGAGAAGTGAGCCCTCTCGGAGTGAGGCCAGGAGTGTGCTGTCTGG
AAGCACAGCCTTGACCTGCCCACCTTGACCTGCCAGGGGTCTGCAGCCTGCTCGCAGGGGTCTGCAGACCGGAGAACAGCAGGGTCACAGGAGTCGTTGGCACCTCACC
CAAGGGGCCCTGCCAGGAAAGCAGGGGTCCTGCAGACTCCTCAGAGGACGAACAGGGTCGAGAACAGGGCTGAGAGCGTGGCCCAATGGTCCC
AAGCAAGAAAGCCATTGGTGAGGTTGGCCCCTAAAGGAGCAGGGAGAGTGGGAAAGGAGGATCCTTCCGTCCAGAGAACCTTAGGAGCCTGGCCCCAATGGTCCC
CAGAAGTGAGCCAGAGTGTGCCCCGGATCACACTGGGCCCCTAAAGGAGCAGGGAGAGTGGGAAAGTCCCAGTGGTCCCAGAAATGGAAAGTCCCAGCCACTAA
AAGCCTATTGACAGGGATCACACTGGGCCCCTAAAGGAGCAGGGAGAGTGGGAAAGTCCCAGTGGTCCCAGAAATGGAAAGTCCCAGCCACTAA
GCTCCTCCTCCAGCAGGGATCACACTAGGTCAGTTGTGAGCCCACCAGTTGCTCTCCTCCCAGCTCCAACCTCTGGAAATCCTTAT
CAGCACTTTCTCCCAGCACTTGGGGACTCCCCAACCTAGTGGCTCACCCGAACAGAAAGGCCAACTTCCAGCCTAAG
TGAGGGCCCAGACAGGGCATCCCCAGAAAGGCAACAGCAACACCCTAGTGCCCCAGACTGGCCCTGATGCCGAGGCTGCC
AACCTGTCTCCCAGGGAGCAGGGAAGACACAGCCACCTGAGCATCAGCCTGATGCCTCAGGTGACCATGCCGAGGCTGCC
TGAGTGTGCACAAGCAGTGAAAGAGGGATCAGCCTGAGCATCAAAAGCATCGGGATTTGGTTCCAGATGAGCTTTAAAG
CCCTCACCCAGAGTCTGAGCCCAGCCCAGCGGGGGCCATCAAAGAACCAGGGGGCCATCAAAGAACCATCTTTATGTCTTTGCTTGCTTGTTTCCTTCGGTCAAC
CAAACATAGCAGTGTTGTTGCCATTTCTTGCACTCAGTACTAGTTGTTGTAATATATGCTCCTGGAACCATCTTTATGTCTTTGCTTGCTTGTTTCCTTCGGTCAAC
CCACATGTAACTAGGTCCTGTGTT

>MASTL_H_SEQID#NA_5
GGTTCTTTCCCGCGGAAGTAGTAGTTGACATTTACAAGGAGCAGCCCCCCAAAGGTCTTTAGCTGTGTTTTTAAGGGGAGAACAGCCTTACCCTCTTTGGACTT
TTTCTTCGTTTTTTTTTTGGAGACGGAGTTGGTCGTTCGTTCTTTCGCCGAAGCGGCGGATCTCGGCGTACAGTGGCGCGGATCGCACTGCAACCTCGCTCCCGGGTTC
AAGCGGATTCTCCTGCCTCAGCCTCGGAGTAGTCGGGGATTACAGGTGCCCGCACCACCACGCCCGGCTGAATTCCTCTTCTACAGACTTTCTACAGCTTCTTATGAA
ATCTTCTGACTGGGCCTTGACTGCAATAAGGTCTTTGCTACAATTTAGTGCTCTTTCTCACACTAAATCAAAACTCTCCCGTTTGGTCCTGATCGTTTCAG
TCAGGCAAATTACATCCTGGGAAAACCTGGAAAAACGTCAGATGACAGGGAGGGCCACTGCGTCTCATCCAGTTGTGGACCACACTTTTGTGTCTTTCATTAGCTTCCAGAC
CTCAGCCCCTGCCCTGCCCTTTACTGTACAGTCAGAACTGGTTTCTACGCCTGCTATGGGAGGTGGGAGGTCGTGTATGGGAGGAGGACCGCGCTTCCCACCAGCCTC

GTTGGGGAAGCCAGGAGGAGAAATCTCTTCAAATCCTGCGATTCAGAGTCAAGTCCCAGTCGTGTCCTTTTCTGGTCGGCCCAGAACTGTTTGTGCCTCCTCCCTCAT
GAGGAATGATGTCAGTGGGCGGGGCCGGGCCGGGTCGCCGCACGAAGGAGTGTAAGGCTGCGAAGTCGGGGGCTTTCCCGACGCCCCCTCCGTCCGCGTCTGCGTAGGG
GAGGTGACGAGGGCGGGGCGGGGCGGGGTGACGTCCAGCGGGTGGGCGGAGCCTCACTTTGAACCCAGTTGGCGGGAGTTGGCTGCTCG
CGGAGGGGCAGTGTCTGCGGGGCCGCGTTAGGCTGTCCAGCGATGGAATTCAGCATAGTGAAGCCCATTAGCGGCGGGGCGCCTTCGGGAAAGTGTATCTGGGGCAG
GCGTGAATAGGATGCGCAGTGCCAAAACCGCCCTCCATTGAGGGAATTCAGCAATGTGTTAAAAAAGCAGACATGATCAACAAAATGTGTCAGTCCAGCTGGCAC
AAAGGCGGCAAATTGTATGCAGTAAGGTTGTTAAAAAAGCAGACATGATCAACAAAATGTCTACTTGGTAATATCTTATTGGGGGAGATGTCAAGTCTCT
TAAGCAAAAGCCCATTCATTGTCATTTGATGAAGAGGATGGCTGTGAAATATATTCTGACTGCACTGGCTCTCAGACTACCTTCACAGACATGGAATCATCCACAG
CCTACACATATATATGTTATTTGATGAAGAGGATGGCTGTGAAATATATTCTTAAAAGTTACTTGGATTTAACACACCA
GGACTTGAAACCGGACAAATATGCTTATTTCTAATGAGGTCATATTAAAACCTAGACAAGATTATTCAAGAAACCCCAGGACAAGTGTTTCTCAAGGACTCGTATGCCTATGTCTGTAGATCAA
ATTGCAGAAAAAAATCAAGACCCTGCAAATTACTAAATCATGTTCAGCGTCTGTGTCTGAAACAGTTGCCTCCACACCTTCAAATCAGTGAAGTGTCTAACTTCTAATTTACTCC
AAGGACACTACGCCTTATTCTAGCAAATTACTCCACATCCAGTGCCAGTAGTCAATCCACAGTTGGAATGCATCAGAATCAGAATGCCACAGCAGTCCAAATGGGAAAA
AGTCTAGGACAAAGGCTGCCACATCCAGTGGGCCAAGCATTGGGCCCAACAATGAGTCAGTTGGAATGCAGTTGAATGCGCAAAAATCTGCAAAATGCTT
GATTGCCAGGAAAGTGATGAAGGATCTGGAGTTAGCTCTTTCTCCCATTCATAACAGCAGTGCCCTTCCCACACTGACACAAGTCAGTCAGTTTCATCAGTGGG
GTTTCAATAAAAAAGGATCTGGAAGCAGTAGAACTGGGATGTAAATATAAATATGGAAGTTAAAAGAAGTTTAAAAGAAGTGTTATACAAAATTATTATACAGA
CTCTGGGAAGTTTCTTGGTGGGATATCTGAAGAGCACCTTGAAGAAGCATAACGAAATGACAAATTTGTTATACAAATCAAAATCAAAATCAACAGGCTTACCTATCCATACCAATGAT
ATAAAAAAACTTGTGTAGAGTATAAGCATAAACGACTGTGCTAATAAGGAGAAATGAGAACATTGTAATTCTTTTACTGATAAACAAACACCAGAAAAATTACCTATCCAAATGAT
GTGCACAAAAGTCAACAAAATGACTGTGCTAATAAGGAGAAATGAGAACATTGTAATTCTTTTACTGATAAACAAACACCAGAAAAATTACCTATCCAAATGAT
AGCAAAAAACCTTATGTGTGAACTCTGATTCATCTTTTCCTGGAATTCTATATGGAAAGTCCATTAGAAAGTGCCTTAGTAGAAGTCAGCCTTAGAAAGTGCAGCATCAAAGA
TAAAAATATTTCTATGAATCAAAATATTGAAGATCCACTTATTGTAACACCAGTTAGCCTCTAAAAGAAATGCCAAGAAAGAAGACCTCACCAAAGGTGTCGAGAACCCTGCGTACAAGA
ATCCTCTCTTTTGAAGAATCAAATATTAGGTCCTCCTTTGGAGGTTGCTGCTGAAAAACGTTAGCCTCTAAAAGAAATGCCAAGAAAGAAGACCTCACCAAAGGTGTCGAGAACCCTGCGTACAAGA
GAGTAACCAAAAATGTTAGGTCCTCCTTTGGAGGTTGCTGAAAAACGTTAGCCTCTAAAAGAAATGGATATTTCGTTGCTTTTGTCATATCCCATGGTCATATCCCTATAACCCTACT
ATCCAATAACTCAGAACCATCCAGAATGAACATGACCCCAAATCAGATCACTTCTTAGATGCAATGGATATTTCGTTGCTTTTGTCATATCCCATGGTCATATCCCTATAACCCTACT
CAAAAAAGAAGATCCTGTATGCCACATCCAGAACCTTGCACCTGAAACACCAGTTGATGGGGCTGTAGATGTGAAGAGTGTGGGCGGGTGGGCCCCG
TTGATGATGGGGCAATTCTAGGAACCCCAGACTACCTTGCACCTGAAACACCAGTTGATGGGGCTGTAGATCTGTAAAAGAGATATCCCTTGGCCAGAAGGTG
GCTTGTTTGAATTTCTAACAGGAAATTCTCAAAGTGCAGTAGAAAATACTTTTAACCATTGATGATACAAAGAGAGCTGAAATGAAAGGCTAAAACGTCATCCTCTC
AAGAAAAGTTATCTGATAATGTCAAAGTGCAGTAGAAAATACTTTTAACCATTGATGATACAAAGAGAGCTGAAATGAAAGGCTAAAACGTCATCCTCTC
TTCAGTGATGTGGACTGGGAAAATCTGCAGCATCGAAAATCTGCAGAATCCTATTTTGAAGCCAGGAATACT
GCTCAGCACCTGACTGTATCTGGATTTAGTCTGTAGCACAAAAATTTCCTTTTAGTCTGTTTAAATGTCTTTAATGTCTTTAATGTGTTTCACGTTTATTGCATAATTATACTCCTTAA
TACTAGATTGATCTAAGGGGGAAAGATCATTATTTAACCTAGTTCAATGTCTTTAAATGTCTTTAAATGGCTGAAAGGAATAT
AGTCAGTAATTTATCTTAACCTCAAAACTGTATATAAAATCTTCAAAGCTTTTTCATTTATTTGTTATTTGTTCACTTTATTGAAGCATCAATAA
AATTAG

>PKC_eta_H_SEQID#NA_6
GAATTCGGACGAGGAGGAGGCAGAAATGGCCAGTGAGGGGCGCTTAGGCGTGGCCTTCCCAGGGCTGCCTCGACTCCTGCACCTGTCCCGAGGGGCTGGCCT
GAATTCGGACGAGGAGGAGGCAGCGGCCCCCCGGGGCGCCCCCGGGGGCGCCCCCACCCGGCGCGTCTGCTGGCACCATGAAGTTCAAATGGCTATTTG
GAGACGGGACTCCCGGTTCTCCCGGGGCAGCAGGGGCGCCCCACCCGGCGCGTCTGCTGACCCCTATCTGACCG
AGGGTCCGCATCGGTGAGGCAGTGGGGCGTGCAGCGCCCCACCCGGCGCGTCTGCTGACCCCTATCTGACCG
GTGAGGCGTGGACCAGGTGCGCGTGGGCCGCGGCCAGGAGCCAACAAACCCACGTACAAGAGACCAACAAACCCACGTACAAGAGGGAGTTTTGGCGCTAACGTCACCGACGGCG

```
GCCACCTCGAGTTGGCCGTCTTCCACGAGACGCCCCTGGGCTACGACCACCTTCGTCGGCCAACTGCACCCTGCAGTTCCAGGAGCTGCTGCGCACGACCGGCG
CCTCGGACACCTTCGAGGGTTGGGTGGATCTCGAGCCAGAGGGGAAAGTATTTGTGGTAATAACCCTTACCGGGGAGTTTCACTGAAGCTACTCTCCAGAGAG
ACGGATCTTCAAACATTTTACCAGGAAGCGCCAAAGGGCTATGCGAAGGCGAGTCCACCAGATCAATGGACACCAAGTTCATGGCCACGTATCTGAGGCAG
CCCACCTACTGCTCTCACTGCAGCCTGTACTTGCCAAAACAATATTAACAAGTGAATTCAGAACAGAGGTTCGGGATCAACATCCCACACAAGTTCAGC
CATCTAATTGTTACAGCCTGTACTTGCCAAAACAATATTAACAAGTGAATTCAGAACAGAGGTTCAGTGTAAAATATGTAAAATGAATGTG
ATCCACAACTACAAAGTGCCAACATTCTGTGCCTCACTGTGGGTAAATGCGGTGGAACTTGCCAAGACCCTGGCAGGGATGGGTCTCCAACCGGACTTGG
CATATTCGATGTCAAGCGAACTCGTTTCCAGATTCATCGGGAACGAACAGGGAAAGGAGAGCAGCAGGTAAAGAAGAAAACAGGAGACCTCTATGCTGTGAAGG
CCAACCTCGAAACTCGGGTGTTCAGGAGGTGTGGGAGGGGAGGGGAGTTTGGGAAGCTTGCAAGGATCCTGTCTCTGCGGCCAATCAACCCCTTCCTCACTCAGT
TATCGACAACTTTGAGTTCATCGAGGATGAGTTGTAAAATGCACCATGACCGAGAAAAGGATCTGTCTCTGGGAGTTGTGAAGATGGAGTTTGTTGATGAA
TGTTCTGCTGCTTTCAGAGATCCTCAGGAAATGACCTCTTGAGGCCATTACCCTACGGCCTGCATTCAGAAGTCTGCGTCGTTTGCAGATATC
GCACGAGCTCGCTTTCTATGCTGCAGAAATCATTTCGGCTCGGCAAGGAGGGGATTTGCAATGGTGTCAATGGGCGTGTTGCTCTATGGCGTCTCGTGTGTTGGAG
ACCACGAGGGTCACTGTAAACTGGCAGACTTGGACCTTACGGGCGAAATGTCTTGAGCCACAGATCCTAAAATCTTTCATG
GCTCCAGAGATCCTCCAGGAAGTGGCCCTGGGCCATACTGAATGATGAGGTGGTCTACCCTACCTGGCTCCATGAAGATGCCACAGGATCCTAAAATCTTTCATG
GCAGAGAAGAAATGGAAGATGAACCCCAACCATGCGGTTGGGCAGCTCAGGGAGGCGAGCACGCCATCTTGACCTGCTGAA
CCAAGAAGAACCCCAAATAGAACCGCCCTTTCAGACCCAGAATCAAATCCCGAGAAGAGTGTCAGTTTTCCTATCGTGTCTCCAGCATCAGCCTTACCTTCAAGGAGAGCAA
AATTGATGAGGGACACATCTTCCAAAGGAATAGAGATTTCCAGAATTCTCTATGGACCCTCCCAGCATCAGCCTTACCTTACCTTCAAGGAGAGTCC
AGGGCACGAGAACCCAAAGAAACTCTGTCGAAGGATGAAGTGAACTTTCAGATATCAACTATTTGAGTCC
GTGAAGAACTCTGTCGAAGGATGAAGTGAACTTTCAGATATCAACTATTTGAGTCC
```

>H19102_H_SEQID#NA_7

```
AAAAAAGATATTTTAAATTTGATGCTCATTTTGTGTGTGTTGTGAGTGCATGCATTCAAATCTGTGTATTCCTCCCTCATCAACCCAGAACATCCTGCAG
CTGCCACTCTGAGGGTGGCCCCTTCCTTCCCTCTGCCTGAAGCTGTATCACCAaAGATTTCTAGTTCCTAGTGTGACTCTGGCCCCATGCTGATGGGTTTCTGCAG
ACTGGAAGAGGCGGGGCTCGTGTCAGCGGCAGCCAGCACACCCGGGTGCCTCACCAAGGGTGGCAACATCCGGGGTCCCTGGG
ATGGGAGCAGTAAGCTGTCGGCAGGGGACCATCAGGTTGGGAAACTCTGGAAGAACTCTGGAAGAACTCTTTCTACCAGAGTTTCCATTAGGCCATTAGG
CCCGAGGCTGGAAGAGCCCAGTACTGGTTAGAGAAGCCTCTGCCAGAGTGGCCAGTGGCCAGCTCCTTTGGAACTGTCTCCAAGGTGCTAGATTGCACCCAGAAAGCTGTATTTGCAGTGAAG
CTAAAGCCAGCCAGCCAGTACTGGTTAGAGAAGATTTTAGGCCTCGTGGCTAAAGCCTCTGCCAAAGGTGCTAGATTGCACCCAGAAAGCTGTATTTGCAGTGAAG
GGGCAGCAGCCAAGGTAAAGGTCCTACAGAGGGAAAACGGCACCTTTCATTATGTGTACTGCAGCACAGATCGTACTCCCTTTGGCTGCTTTCCTAGAT
GTGGTGCCCAAGGTCAGCTGGCAGGGAAAACGGCACCTTTCATTATGTGTACTGCAGCACAGATCGTACTCCCTTTGGCTGCTTTCCTAGAT
GGGGGACAGCTGCAGGGAAAACTGGCAAAGTTGTGGTACTGCTGTTATCTCCATGACTTGGGCATCATGAGCTCAGTACATGGCTGCAGAG
TTCCATCCGTGTCTTTGTGCCGAGTTGGTGCTGTTATCTCCGCCAGGTGGTCCCGTGATTGGTGGTGCTCTTGCTTTCTCTGTGGAAAGTTTCCAGTGGCTGCAGAG
GAGGTCCTAAGTGACATCGAAACTGAAGTGGAGGACCTTACAACCATGTGACCACAGTGTGACCCAACAGTGTGACTGGTGCAAGTGCAACTGAGCTCTTATGCC
AGAGATCATGTGGCCATGTGGCAAGTGTTATCTGCGTTATCTGATCACTTCCACCTTTCTTCGGGGTGTGCCTTGACCACCAGTGTGACCAGTGAA
AGAACCCCCTCCAATTCGTCTACGTTATCTGCACCCAGCCCAGTTCAGCGGAGACCATGCCCTTCAGCGGAGACCATGCCCTTGACTGTGATCTGGAGTCTTGTGCTCTCACCCTATCCCT
CTTTGTCACGGAGACACAAGCTACCCAGCCCAGTTCAGCGGAGACCATGCCCTTCAGCGGAGACCATGCCCTTGACTGTGATCTGGAGTCTTGTGCTCTCACCCTATCCCT
GCTTGAGCCTCTACTGTAAATTGGGCCCGG
```

>MSK1_H_SEQID#NA_8

```
TGCGGGCCAGAGCGGGGAAAGAGACTCGTCTTTGCGTCCGAGTTCTGGAGCCGCCGCCGCGCACCCGACTCCTGGGCCGCCAGCGGCCTGCGAGGGGACGGGGCGT
CCGCTGTCTCCTGGGTTCCCCTCGTAGCGACCCGCGGATCGGAAAAAAGGAGAAGATGGAGGAGGTGGCAGCAGCGGCCGCCGGGGGACCCGGGGGGACCA
GCGCGGACGCGGCGGGCGACGGAGGAGCAGCTCCTCACTGTCAGCTTATGGAACTGGAGCTTATGGAAAAGTATTTCGTGTAAAATAAGTGGCCATGAAGCTGTATGCCATGAAAGT
TTTTGAGCTCCTGAAGGTCCTAGGAACTGGAGCTTATGGAACCACAGACAGCATACAAGGACACAGTCCTGGAACACATTAGGCAGTCGCCATTTTGGTAA
TTTGAAAAAGGCAACAATCGTTCAAAGGCCAAAACTTCACAGAGAAACCAAACTTCATCTCATTTTAGATTATATAAATGGTGGTGAACTTTTCTCAAAGAGAGGCGTTTCAC
CATTACATTATGCTTTCCAGACAGAAACCAAACTTCATCTCATTTTAGATTATATAAATGGTGGGATTATAAACTGAAAGAGCATATTCCTTTGTGATTAAGCTTGAGAATATTCT
AGAGCATGAGGTGCAGATTTATGTTGGAGAGATTGGTCTGCTGCCCTCGAACATCTCCACAAGTTGGGATTATAAACTGAAAGAGCATATTCCTTTGTGATTAAGCTTGAGAATTAAT GAA
ACTTGATTCTAATGGCCATGTGGTGCTGACAGATTTGGTCTGCTGCCACAGATTTGGTGGCTGATGAAGAGCAGTTGACTGGTGGAGTTTGGCTGATGAAGAAATTACTAACTGAGC
TACATGGCACCAGATATTGTCAGAGGGGGAGATTCAGGACAGACAAGGCAGTTGACTGGTGGAGTTTGGTCCACGTGACTGAGATGAATTAAAAGTGAGCCTCCATATCCCAAGAAAATGAGTGCTTTAGC
ATCTCCTTTCACTGTTGATGGAGAAAATTCCCAAGTCCAAGAAGAGATTGGTGGTCCAGATATCTAGGAGAATATTAAAAGTGAGCCTCCATATCCCAAGAAACATCTCTCTTTCAGA
GAAAGACCTAATTCAGCGTCTTTGATGAAGAATCCCAAGAAAAAGTGCCTGCCCAGAGTTCTCGAAGCGCTAAAATGTTGCAGGAGGTAACTTTGCAGAAGAGTTCACA
GAAAATAAATTGGGATCCCACTTATTCTCCCGCGCCAAAAAAGTGCCTGCCCCAGAGTTCTGAGCGTTCAGGCTATTCCTTGTTGCCTCCATCCTCATCCTCATCAAC
CAGCTGTGTCATAGACCCCTCTCAGTTTCACATGGGAGTTGAAGTGAACGTCCTGGAGTTGGAAGGTAGTTTGTCATTGTGTGTTCAATTTGTCGAAGTGTGTTGCATGAAGTTTTCAA
ACTATGACCTAGATTTGAAGGACAACCCCTCTCAAAGGGAGTTGAAGTGAACGTCCTGGAGTTGGAAGGTAGTTAATTGTGTGAAGCTCTGTGAAGGACACCCCAATATTGTGAAGTTGCATGGAAGCCAGCT
ATAATCAGCAGCTTCACCGTTTCTAGTGAACTGTTTGAATGGAGGAAACTGTTTGAGTGGTGTCAAGCCACTGTGTGTGAAGGAGGAGCATTAAGAAAAGCACTTCAGTGAGACGGAAGCCAGCT
TGATCAGTTCACACGTTCTTTCAGTGAACTGTTTGAATGGAGGAAACTGTTTGAGTGGTGTCAAAGCCATTAAGAAAAGCACTTCAGTGAGACGGAAGCCAGCT
ACATCATGAGGAAGCTGTTTCAGCTGTAAGCCACGTGTGTGAAGCGGCTAAAGCCACCGGATAATCAGCCCGTCATTTGTACACAGTGTTGTGTCAGGACAGGT TCCTTCAATCTCAT
ACAATTTGGAAATTAAAATAATTGATTTGGATTTGGTGAGCTTGTGTGAAGCGGCTAAAGCCACCGGATAATCAGCCCGTCATTTGTACACAGTGTTGTGTCAGGACAGGTTCCTTCAATCTCAT
AGAGCTCTGAATCAGAAACGGCTACGAGGCGCGGTGGAAATCATGAGAAAATTAAAAAGGCTTAAAATGTCTGGCTTGAGGTACAAGATGGAAGTCAGC
GACCGAAGTTGACGTGTACCAGGACTTGACCTGTACGAGGCGCGGTGGAAATCATGAGAAAATTAAAAAGGCTTAAAATGTCTGGCTTGAGGTACAAGATGGAAGTCAGC
GGCTAAAGATTTGATCCAAGGACTTCTCAGAAGACTCCGGATATTCTAGGAGATCTTCCGGAGCTGCCGTGCATACCTGTGTGAAAGCAACCTTAACAAATACAAGAG
TGTCCTCCAATCTCTGATGACTCCGGATATTCTAGGAGATCTTCCGGAGCTGCCGTGCATACCTGTGTGAAAGAACTAGCACCAGTACCGGACGCGCAGTTCCAGTG
AGAGGGGTTTGCCTTCAGAATGTTGATAAGGCCCCTTTGGCTAAGAAGAAAATGAAAAAGACTAGCACCAGTACCGGACGCGCAGTTCCAGTG
AGAGTTCCCATTCTTCTCCTCCATCTCACGGTAAAACTACACCCACCAGTGATCCTTTATCCCTGTTGGTTATCTCATTCAAAGAGAAATGTT
TCCAGTTCTCGGACTCAGTAGCTTAGGCATGGAATTGCCTCCTAATGATTTTTTCAGGAAACCTGTTTGGTTATCTCATTCAAAGACTGGACAGAGAAATGTT
AAAAATGTTCCCGTTGGTCATTGGAATCTTTTTAGCAACTGGAGTGTTTTTCAGGAAACCAAGAACTATTCTCTGTAAATTTAATTAGGGCTAG
ACTGTGAATAGAGCACATTACTCTTTTGCCATTTTTGCACTGTTTTTAGCAACTCTCTTTCTGTAAATTTATAAAGTATGAAATAAAGCAAACTATTTATAAGTGCACTG
ATTTGAGCTGCTTGTAAGTCACAGTTTTCCAGGTTTCCAGGTTTCCAGATTTATCAGAGAAAGGGTGCTGTTTTCAAAGAAGATTCTATGTGTCTGTGTTTGCCTTGTGTGGCATATACTTATTTCTTG
TTTGAAATTTGCACCCTTCAAAACAATGATTTATCAGAGAAAGGGTGCTGTTTTCAAAGAAGATTCTATGTGTCAAGTGAAAAAAAAAAGATATA
AGAGAAGAGATTTTAACTTAATTGTTTTATTTTATGTTTACATAGTGATAAACTTTTCTAAAAGTGAAAAAAAAAAGATATA
AGAACTCAAGGTCCCATACTCTGTATTCGGGAATCCATCTGAGATGCATGCTAAGCTATGTGTTTTAATTTGCACTGTATGTTTTAATTTGCACTTTATAAGTGCACTG
AATGGTTATTGCAGAATATTAAGGTACACATGTCTCTGTTTTAAGTAATTGCACTTTATAAGTAATAAAGCAAACTATTTATAAGTGCACTG
TTTAAAGCATTTGCACTGTATTTTGCCATTTTTCATTTTCTATTTTCATTGTCCTTCACTTTTGTATGCAACAGTAGAATGGGGCAT
GTTGTGTCATGTAGTCAGCCACTTATGCCACCAATGTGAGGGAAATTAAACTAAACACTGTGTTTGTACACTGTTGTACTAC
AGTGAGGAAATTTCCTCCTGTAGTCATAATATTTAGAATCATAATATTTTAGAATCATATGACTTGTTTGGAAATTTTCTGTTAATTTAAATCCAGAAAG
CATATTTTATAAACTTATGCAGAGCACTTTATTGCTCAAAAGTTCTGAAATTCATGATGAAAACAAGTACTATTGCAATTTCATTCATTAAAGATTGCT
GCATTTAAAAATACACAATTAATTCGTTCCCTATGC
```

>YANK3_H_SEQID#NA_9
CTGCCGCTCCCGGCCTGCCGGCCGGCCATGCCCGGCACCCAGACCCGGCACCCAGGGCAGGGCAGCAG
CGCCGCGGCCTGCCCGGCGTCCCCGGCGTCCGCGCTCCGACCGGCCCTGCCGCCCGTGCTGGCGGCCACCGCGCGGGCGGA
CCGGGACTCGGCGATGTCCGATCCGGCCTCGCACCTTCGACCGTGAACTTGACCACTTCCAGATCCTTCGGGCCATTGGCAAGGTGTCATTGTGCAGAAG
GGCCGGTGTTTGACGACAAGGAGGAGGAGACGTGAACAAGCAGCAGTACATGAACAAGCAGCAGTCCGCAACGTCTTCGGGAGCTGGAGATCCTGC
GGGGACACGGAGAAGATGTACGCCATGAAGCAGTCCATGCGGCGACAGTCATGGGCGGCAACCTGGT
AGGAGATCGAGCAACGTGCAGTTCTTCCGGTGAACCTGTTACTCCTTCCAGGACGGGCACATGCTCATCTGGCCTCTGGGCGCGGCACCTGCGCT
ACCACCTGCAGCAGAACGTGCAGTTCTTCCGGTGAACCTGGTTACATCTGCAGACAGTGCAGACAAGCCGGCGTACATTGCCACCTCAAACATTGCCCAGGCTCAACTGCAGCACAATCATC
CACAGAGATGTCAAGCCTGACAAGCCGTACCAGGCCTATGCCGTCACAGGCCTGATGAGGAGACAGACACTGGCGGTCAAGGACGGGGAGCGGGGC
GACGGGCATTAGCAGGCACCAAGCCGTACAGATGGCTCCGGAGATCTTCCACTCTTTGTCAACCCGGGGACTGGACTGGTGGTCGGT
GGGGGTGAGGCCTATGAGTCTGGGAGGCCTATGACTGGCGCTGCTGGCCAAGGAGAATGGTGGCCCTGCTGCAGCTCTCCTGTCAGTCTCCAGGACGTGCAGG
TCCAGTATGTGCCCAGGGCCTTGGCTCTCACGTGGCGCCGGTGCTGATCCTGCAGGGCCCCCTGAGCCGAGGTGAGCACGGTGCTGTGCGCGCCACCCCACC
CAGCCCCGGCGCTGGCCGGCGTGCTGAGTCCCAGGCCCCTGCACCTGGAGTCCTCACTGTGAACCCGGCGTCGCTGGAGGTGGAGAAGCGCCTTCGTTGCCCAACAAGCCGTTCGTGCCCAACAAGAGAGAGACGAGTTCCCGAGG
TTTGAGCTGGAGGGAGTGATCATCTTCAAGGAGCTGGCCTTCAAGGATCCTGCGGATCTGTGGGACGTGTTTAACGAGAAGCTTCGTGATTTTAAACAGAGAAAAGCTGAAGACTTGCCCCTGCGGCC
AGTCCGAGAAATGACTATCTTCAAGGATCTGGAGTTCAGGGATGCTCGAGGCCTGGAGGATGCTGCGCCCCTGGTCTCCACCCGTGCCCCATTTCGCCCTGTGTGGCATCACG
GAGCCTCTCCCGCCCCTGAGTCCAGGGGATGCTCGCCGAGCGCTGGAGGCCGGAGAGGCCTGCGAGAGGGCCGAGGCCTGGCATTCACG
GGGAGCGGCTAGGCCGGGACAGCGCCCGTGGTCCCCACCAGTGCCCCGGACACCTTCACACCCGACACCCAGCTGCCCCGGGTGGGCTGTGCCAGGGGA
TTCCCACCCAGCCTGGCCTGGCCTGGCGTGGCGCGGCTGCCCGTGGCCCCCGGAGACCGGCTAGGCCTGCTTCCCGTGGGCCCCTGGGCTGGAGAGGTCGTGCAGGGGAGCTGTGCGGGGAGCTGGGGGCCATGGA
CACCTGTGGAGGGCCATTCCCGTGCCCTGCTCAGCCCGGAGGTGGCGGCCCTCGTGGTGCCACAAGGTGACCCCTCGTCAGGCAGGGAGCTGGGGGCCATGGA
TGGGACAGGAGTCTTTGTCCCTGCCTCAGCGCAGGCCCGGTGTGCACAAGGTGACCCTCGTCAGCACAGGCCGCGGGTGCCCAGCTCGGCTC
AGTTCTTGGAGGTCAAGGGCATGGGTTGGGGCTCTGGGGTGGGGTAGTGGGGTTGGGGAGGTTGGGGGTAGTGGGGTAGTGGGGTTGGGGGTAGTGGGGTTGTGGGATAGTTTTCACCTC
TGAGAATTACAATGTGAGAACCGCTC

>MARK2_H_SEQID#NA_10
CGGCGGTGGTGGGCGGCCATGTTGGGAGCAGCAGGTCCGGCGGCCTGCCTGGTGCCGGCGGGAGGGCAGGGGAGGAGGAGAGGAGAGGCGAGGC
AGGCGGCCGGGCTGCGGGGCGGCAGAGAGTAGGCGGCGGCCGCGGGGGGGTGCGGGTCGGGAGCC
GCTCGGAGCCGCGGCGGCCTAGCCCGAGCGCGGCCGGCGGCATCCCGGCGCTGCCCGGGCCATGCGGCGGGGT
GCTCCTGCGTGTGAGAAGGCCGGGCCTTCGGTTCCCGGCCTTCCTCGGTTCCCTCCTTCCCGGTTCCCTCCGGTTCCCTGACTGCACTGGAGGGCCCCACCTGGGACACTGCAGAGGGCAGCAGGGACACTGCCTGGAATACGGCGGCCATGTC
CAGCGGCTCGGACCCCCTACCACCGCCCCTACTGCTGATCAGCGAGGCCAACGGAGGACACGGACACGGAGGCAGCGCCCACTTGGGACACCTTCAAGGGTCAGGGTAATTTTGCCAAGGTGAAGTTG
GGGGCCGGCAACTCGGATCTCAGCAGCTCGATGAGGAGGTAGCGTGAAGAGGAGCAGCCCACCAGGGCCAGGGTAATTTTGCCAAGGTGAAGTTG
GCCCGACACATCCGACTGGGAAAGAGAAGGTAGCTGTAAATTATTGAAGTGATTGAGACTCAGGGCAAACCTCCAGAAACTATTCCGGAAGTAAGAAT
AATGAAGGTTTTGAATCATCCCACATAGTTAAATTATTGAAGTGATTGAGACTCAGGGCAAACATTGATGAACCCGATAGTGTCATGGGGGAGAGGT
ATTTGATTACCTAGTGGCTCATGTGCTGCTTCAGCAATGGACTTGGCTTCAGCATCAAGATTGCAGACATCAAGAATATGATGGACCGAAAAAATATGATGTTGTGGGAGGAGTTTAT
TGTTCCATAGAGACTTAAAGGCAGAAAACCTGCTTCTTGGATGCTGAAGATCTGTTCAGGGGCAAAAATATGGGCAAAGGCTAGGAGTTAT
CCTCTATACACTGGTCAGCGGATCCCTGCCTTTGATGGACGAACCTCAAGGAACCTCAAGGACAGAAATAACCGTATTCCATTCTACAT
GTCCACGGACTGTGAAAACCTGCTTAAGAAATTTCTCATTCTTAATCCCAGCAAGAGAGGCCACTTTAGAGCAAAATCATGAAAGATCGATGATGGAATGTGG

GTCACGCAAGAGATGATGAACTAAAGCCTTACGTGGAGCCACTCCCTGACTACACAAGGACCCCGGCGGACAGAGCTGATGGTGTCCATGGGTTATACACGGGAA
GAGATCCAGGACTGCTGGTGGGCCAGAGATACAACGAGGTGATGGCCACCTATCTGCTCTGGGCTACAAGAGCTCCAGCTGGAAGGCGACACCATCAC
CCTGAAACCCCGGCCTTCAGCTGATCTAACCAATAGCAGCGCCCAATTCCCATCCACAAGGTACAGCGAAGCGGTGTCGGCAAATCCAAGCGCGGCGCT
TCAGCGACCAGGCTGGTCCTGCCATTCCCACCTCTAATTCTTACTCTAAGAGACTCAGAGTAACAACGCAGAAAATAAGCGGCCTGAGGAGGACCGGGAG
TCAGGGCGGAAAGCCAGCAGCAGCAGGAATTCCCCACTTTGAGCGGGGCCCGGCCCAGCACGCACCCCAGAATGCAAAGACAGCCTAACCATGCCAG
CTCCACCAGCACAAATGCAAGCAGGAATCAGGCCTTCTGCTTCTGCGCGCAGTCTTGCGCGGCCCCAGCACGCACCCCAGCACGTGCCTCCCATCCGCCCCACAAC
GGTCCCGGCCTCCACGGCTTCTGCTTCTGCGCGCAGTCTTGCGCGGCCCCAGCACGCACCCCAGCACGTGCCTCCCATCCGCCCCACAACATCAGCAGCA
CTGGGCTGCCCCCACGGAGTAACTGTGAGGTAACTGTGAGGTAACTAACTTCCCGGGTGTGTCAGCCAGGCCGGCGGGGTGTCACCTCCAAGTTTGTACGCAGA
GTGGTGGAGCCCAGACCGAACTAACTTCCCCGCTGCCCAATGAACCTGAAAGCAAAGGACCGAAGCAAGCAGCCAACTGAGCAGTGGGCAGTGGCGGCAACGA
CCCTACGGTGTGACCCCAGCCCTCTGCCCAGAAGGAACCTGAATGAACCTGAAAGCAAAGGAAGCAAAGGACCCCGTGCCCAAGGAGCCCAACGAGATGATGCGG
AATCTGTCTTTCAGGTTTGCCAGAAGGAAGAATTTCGGGACGCGAACAGCTGCCAAGCGAGCGCAACAGCTGCCATGGAGAAGTACATGTCGTGTGCATGCAGAGGACTTCGT
CAAAGAAAAGGAAGAATTTCGGGACGCGAACAGCTGCCAAGCGAGCGCAACAGCTGCCATGGAGAAGTACATGTCGTGTGCATGCGGATATCGGGGCCAGCTGGGGGCCCACCTC
GAGATCCGCAAGGTGCTGAGGCGCAACAGCTGCCAAGCTGCAAACTGCGCGGCTCTCTCAAACGGGTTCGATTTAAGCGGATATCGGGGCCAGCTGGGGGCCCACCTC
GCAGTGGGAGATGGAGGTGTCAAACTGCGCGGCTCTCTCAAGACGGGTTCGATTTAAGCGGATATCGGGGCCAGCTGGGGGCCCACCTG
CAAAATAGCCAACGAGCTGAAGCTTAACAGGCTGGTGTCTCCGCTGGCACTTTCTCCCACCATTCACCCCTGCGCGCGCGCCCAGATTCCCC
GGCGAGACTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
CTTCTCCTCCCCTACTGCGAGCAAAGAGGAAAGGGAGGAGGGTGGATGGGGGGGCAGGGGCTCAACCCTCGCCCCCCTCCCCCCCTGCACCAGAGTAGTATTTCGCTGCTAAACCA
AGAAATTTTTATTACCAAAAGA

>NuaK2_H_SEQID#NA_11
GGGCGCGCCGGGGGGCGTGTCCAGGGCCTCGGGCTCAGGGGGCCGTTTATAACAACTTGGAATGCTGTGCTTTACTGCGCGCTCTGGTACTGCTGTGGCTCCCCGTC
CTGGTGCGGGACCTCGGGAGCTTCAGCCTGCTCCCCGACAGCCCTTGCTCACCTCGTGCTCCGCCATGGAGTCGCTGGTTTT
CGCGCGGGCGGCTCCGCGCCCCACTCCCGGCCAGAGCTAGCCCGGCCGGCTGGCGGAAGGGCTGATCAAGTCGCCCAGCCCTAATGAAGAAGCAGGCGG
TGAAGCGGCACCACCACAAGCACAACCTGCGGCCTACGGAGTTCCTGGAGACCTACGAGTTCCTGGAGAGAAATCAAGATGAGCAAGATCGATGCACATACGGAGGGGAGATTGAGATCATGTCATCAC
CTCGGGGGCGCGCTGGTGCGCCATCAAGTCAATCCGGAAGGACAAAATCAAAGATGTGTTTGAGAACAGCAGCAAGATGTCAGCCGGGGCGACCTTTATGACTACA
TCAACCACCCTCACATCATTGCCCATCCATGCGGCGAAGCTCAGTGAGCGGCGAAGCTAGGCATTTCTCCGGCCAGATATCAAGATTGCTGCTCTACCCTCAAGGCAAGTTCCTGCAGACAT
ATCTCAAGCTGGAGAACATCTCTTGGATGGCCTCCGCCAGAGATTGTCAATGGGAATATCAAGATTGTCAATGGGAAGCCCTACACCCTCTGATGCC
TCTGTGGGGAGCCCCCTTCTATGCCTCGCCAGAGCATGGCATGAAGCTAGTGAACCATAAGATCCTAGTGACCATAAGATCTAGTGACCATAAGATCTAGTGAAGCCCTACACCCTCTGATGCC
TGGTGCATGGCACCCATGCCCTTTGATGGCTGTTGATGGTGAACCCGGGCCAGTGACTCTGCCCGGGGAAGCCACCCTGGGGTCAACTGCTCACCGCGTTCCTCCCGCCCCCTCCT
AGTGGGAGAGCAGGAGGCTCGAAAGGTGCCAAGGGTGTCGCTGCAGCTCTTCAAGGCTCGCAGCTGTGGTGGCTCTGCCCGCGTTCCTCCCGCCCCCTCCT
GGAGAATGGGGCCAAGGTGTCCACAGTGTCGCCAGCGTCTTCAAGCATGACAGGGCCCCTGGGGAGCGCCAGCATTCGCCAAGAGTGCCTCAAGGGCATTCTC
GCAAGGAGAATGACATGGCCCAGTGTTGCACAGTGACAGGGGTACAGGAGGTACAGGAGGGTACAGGAGCCCCTGCTCCCAA
AAGAAGAAGAAGGTGTCAGCCTGCTGCAGAAGGGTACAGCAGCCCGACACCCCCGGGGCGCAGGGCGACGTGTTTGT
GAAGGGCATTCTCAAGAGATCCGGGGATCCCAAGAGAGCACCTCCGCAGAGCCTCCGGCTCTCGGCTCTCGGCTCTCGGCTCTCCCGAGACAGCCT
GAGTAGCTCGGCATCTCCCTGAGTTCCGCTCCCCTGATGAATCGGGGCCAGCCGACCCTGGGGGCTGTGAGCGGGAGGACA
GCATCCGTCCTCTCTGAGTCCTTTGACCAGCTGACTTGCCCTGAACCGGCTCCCAGAGCGGCTCCCAGAGCTCCCAGAGCTGGGGGGCTTG

AGGAGCCCCCTCAGAGGGCCCTGGAAGCTGCCTGAGGGCGGCTGGGCGGCAGGATCCTTGGGGGACAGCAGCTGCTTTCCTGACAGACTGCCAGGAGGTGACA
GCGACCTACCGACAGCAGGCACTGAGGGTCTGCTCAAAGCTCACCTGAGTGGAGTAGGCAITGCCCGGTCCCAGCTCTCAGATGCAGCTGGTTGCACCC
CGAGGGGAGAATGCCTTCTCCCCACCTGGGACCTGCCATCCCAGACCTCAGAAGGCTGGAGGGGTTTGCAGTGGAGCCCTGAGCAGGGAATGCCCGCCACAGAGT
GTAGGCAAATGAAATGCGCCAAGGGTTCAGTGTCTTGTCTTCAGAGACATCTCCCAGGGCCATCCTCTGCAAGTGGGGCCCATCACCACACTCTGTACACAT
CCACATTGCCTGTTTCTTGTTGTACATGGAGGGGCCACAGAGACCTGGAAAGAGAACTCTCCCAGGGCCCTCTTTACACACACAITCCCGTTCCTACCACCAGAACTGGAT
GGTGCCTTCTAAGGACAGCTCCTTCCCTACTCATTCCCTGCCCAAGTGGGGCCCAGACCCTCTTTATTTTTATTTTTATTTTTATTTTTTGAGACGGA
GGTGGCACCCCTAAITGTGCCAGCTAGGAGTGCAATGGCGATCTGGAAATGGTCTGGAGTAAGCTTCGTGGTCTGGAGTGGCATCCTCAGCGATCCTCCTAG
GTTTCGCTCTTGGTGCCGCCCGCTAGGAGTGCAATGGCGATCTCACCTGAGTCCTCCCGGGTTCTCCATGTGGTCAGCTGTCTCAAACTCCGACC
TAGCTGGGATTACAGGCGCCCGTACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCAACCTTCCTTATTTAGCTGGGAGTAAGGAAGAA
TCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCCCAACCTTCCTTATTTAGCTAGGAGTAAGGAAGAA
CACAATCTCTGTTTCTTCTTCAATGGTTCTTCCCTTTCCTTCAAGGGGCCCAGCCTCCTGGAGTGGCCTGAGCTGTCTCAAAGTCC
CCGCCCTGCTGTGTGGACTTCATCTCAGGGGCCCAGCCTCCTGGAGACTCCACCTTGGACTCCAGTTATTTTTATGTTCTTGGTCAGAACTTCTGCCTCTAAAGTCC
AGACTATGGAATGCTATGTTCTGGGAAAATCCACTCAAGTTTTGTGTACAGAGAGATATTTTGCAACTATTTCCACCTCCTCCCACACCCCCA
CACTCCACTCCACACTCTTGAGTCTCTTTACCTAATGGTCTCTTACCTAATGGTCTCTTACCTAATGGTGCCAAAAGTACCATTAAAACCAGAAAAGGTGATTGATTGAAA

>BRSK2_H_SEQID#NA_12
TGGCCGGCGGCTGGCCCCAGAGAGCGATGACATCGAGCGGAAGGACGGCGGCGGGGAAGGACGGGGCAGTATGTTGGCCCTACCGGCTGGAGAAGAGCGCTGG
GCAAGGGGCAGACAGGTCTGGTGAAGCTGGGGGTTCACTGCCGTCACCTGCCGTCAACCGTGGCCATCAAGTCGACACCCCCACGTCTCCAAAGTCTCCTGAGAAGCTCACCGTGAGAAGCTCAGCGAGTCGGTG
CTGATGAAGGTGAGCGGGAGAATCGCGATTCTGAAGCTCTGAAGCTCTGAAGCTCTGAAACAAAAATATTTGTACCT
GGTGCTAGAAACACGTGTCAGGTGGTGAGCGTGAGCGTCAGGTGGTGAGCTCGGAGAGTCTTCCGGCAGATCATCT
CTGCGCGTGGACTTCTGCCACAGCCACTCCATATGCCACAGCCACTCCATATGCCACAGGGAGAAGAACAACATCCGCATGCAGACTTTG
GCATGGCGTCCCTGCAGGTTGGCGACCAGCTGTTGGCGACCAGCTGTTGTGGTCCCGAGTGATCCGGGGGAGAAGTATGACGGC
CGGAAGGCGGACGTGTGAGGCGTGTGTTCCACATGCGCGCACTTTATCCGCGCACTTTATCCGGACTGCGACAGTCTGCTGTGGAGAAGGT
GAAGCGGGGCGGTGTTCCACATATGTATATAGGGGGCAAGAATGAGCCGAACCAGAGACCGAACCAGAGACCGAAGGCCAGCTAG
AGCACAITCAGAAACACATATGTGCTACGTGCTGACAGCATGCACGTCCTGGACAGCATGCACGTCCTGGAGGAGGAGAACCA
GGAGAAGATGATTTACTTCCTGGACGAAAGAAAGTACCCGGAAAGAAAGTACCCGGAAGCGTGACGGACCTCCCGGTGCCT
AGCGGTGTGGACTCCCGATGCTGAAACGGGCAAGGCGAAGGCGAAGGCGACCGGCTCTCAGCAGCCCTCCGGTGCCT
GCCGGCGGCCAGTTGAGATGGCCCAGCACGGGTCTCGGTCGGGTCTCGGTCGCACGCCCACTCAGCAGCCCCAACCCCAGGGT
GACCCCTCACCCTCACCAGCGTCGGAGGGGCAGTCCCACGCGTGCCACTCGGAGGGGCACACGCGTGCCACTCGGAGGGGCGGTGGCCAGCAGGCCCGGGT
CCCCGTCCAGCGCCGGAGGAGATGTCCAACCTGACCACTCCATCAAGAACAGCTTTCTGGGCTCACCCGCGGAAACTGC
AAGTTCCGACGCCGGAGGAGATTCGTGGTCATCAGAAGACAAAACCTCGAGCTGCTGCACTCCATCCGGAGATCATCAGCTGGCACTTCATCAGTCTCAGCAGCCTGGAGAGGAG
GAGCAGATCTTCGTGGTCATCAAAGACAAAACCTCGAGCTGACATGTGGCACCGTGTGTCGTCAGTTCCAGTTCCAGTTCCAGGCACAGCGTCATC
TCCCAAAACGAGCTTCCGGGCCGAGTACAGGCCCAGCCGGGGGCCAGGCCCAGGGGGGCCGTCAAGTTCCAGGTTGATATCACTACACTACGGAGGGGTGG
GGAGGCGCAGAAGGAGAACGGCAGCATCTACGTGCACCTTCACCTGTCTCACCCTGTCTCACCCTGTCTCAAGAGGGTGGTGGAGACCATCCAGGCCCAGC
TGCTGAGCACACACGACCCGCCTGCGGCGAATCCAGGCCCAGCCTGCAGACCACTAACTGTATGGAAATGAGCAGGACTAAGCTGGGGGCGGCTTAAGGGCCAGAGGTG
GCCAAGCGAGAGTAGGCCCTGACGCTGCAGACACCACTAACTGTATGGAAATGTGAAATTATCCGAAAAGT
TAACATGTCACCTCCACGAGGCCATCCTCTGTGACCAGCAGCTGCTGCGGAGCGTGCTGCGGAGCGTGCTGCGGGCCAGTGAGGCCCAG

CCCAGCGCGCCGTCCACCCGGCAGCTCCTGCCTCAGCTCCGCACGGCCCGTGGGAGGAAGGCCAGGCTCGGGGAGCCTCCTCCAGCCGGCCGGACC
CGGACTCCCGGTCACCTGACCCCTCAGCAAGAACAGCCTGGTGGCCTTCTGGGGCCAGGACCCCGGTGGGCAACGTAGCCGGGAACAGAGCCCCG
TCCACCGCGCTCCAACGCCGCACCCGCCCTGGAGGCCTCCTGCCGCCCTCCCTTGCCTCCATCTGGGGCGGCTGTGGCGGCTCCGTGTAGTCTTGGCCTCCTCAGGCTGCCT
CCCGTCCTCTCGTCTCAGCCAGGCGCCTCCTTGGCGGCCTCATCTGGGCGCGCTGTGGGCGGCTGTGGGGCTCGTGGCGCTCCGGGAGGTGGAAAACAGAGACACCCTGTGGCA
CCAGAGCCTTCCCAGCAGGCCAGGCGCCGGTCTGGGCATGGGGGATCCAGTGTTTATTTGGCCGTTTTATTTGCCGGGCCCTCTGTTCAGGGAAGGGGCGGC
GGCCACATCCCCTGCCCTGCCGCGTCTCAGGCAGTCAGGTGGGGGGGCTGGGGGCCAGGGCGCCCCTCTGGGCGCGGGGGGCTGGGCCAGG
CTACTGTAAACTTTAAAGAATTCCTGCACGAATATTTTTATAAACTTTTTTCTGTGGTTTTGAGAAAAGGTGTGGGGTGTGGGGCCCGTGGGCAGG
GCCAGGTTTTGTTTTAGTTCCCCTTGCCCTCCTCTTCTTCTTCTCACACACACACCTCAAAGACGGTGCGCTGCTGCTCGTCGTTCCGTCGTGGTTTTGT
GTCCAAAGGTGAAGGGTAGGAGGGTAGGAGAGTCGGTTGCGCTCGTCCAGGGGCGTGTCTAGCTGGCCCTGCAGTGGCTTTCTTATTAAAATGAAAAATT
GAAAAAAAGGACAAAGAGTCGGTTGCGCTCCTCTGCAGGGGGTTCGTGCAGGGCGAGGCAGGGGCAGCCTCAGGAGGGCTGCAGCGCGGTGCGGCGGCCG
CCAGTGCCCGCCGTCGCTTCACCCAGCTCAGTGCCTGCCAGCTCCAGCTTGTGTTCCTTGTCTGTCTTCGTGGCCTCCACCACCTGTGTAGATTTTAACGCTT
TGGGGCGGGCGGGGGTCTGATGCATGCCACAGCTGGGATTTCTATACATAAGAACAAAAGCACACCTAGGACAGAAACGCCAGGCGGTACAGGCGGGAAGGG
CTGTTAACATTAGACCTCTGCCACAGCTGGGATTTCTATACATAAGAACAAAAGCACACCTAGGACAGAAACGCCAGGCGGTACAGGCGGGAAGGG
GCTCCACGAGGATCGAGGACACGAAGCAAACTGCCTCTTGCTTGCCCTTTTGTGCCTTCCCGTTTTGTGCTTGGGACACACGCGACTCCAGCAGGCGCCACGGAAATGG
GCAAGCCCCTGCAGTGTACCCCTGTCATAACTGTGAGCAGCTGCAGTCGTCAGCTCAGCAGCAATAAATCCCTTCCGCAAAGACAA

>MARK4_H_SEQID#NA_13
CCGGGACCCGGAGAAGAGATGTCTTCGGCGACGGTGCTGGCCCCGGGCAACGATCGGAACTCGGACACGCCATGGCACCTTGGGCAGTGGCCGCTCCTCGGACA
AAGGCCCCGTCCGGTCCAGCCGGTCACTGGGTGCCGGAACTCCATGCCTCCTGTCCGGGAGGCAGCCCCCACGTGGGCAACTACCGCCTGCTGA
GGACCATTGGGAAGGGCAACTTTGCCAAAGTCAAGCTGGCTCGCGGCCACATCCTCACTGGTCGGGGAGGTTGCCATCAAGATTATCGACAAACCCAGCTGAAT
CCCAGCAGCTGGTGTGAGGATGGAGGTCTGCAGAAGTCGTTCTCCGAGAAGTCGCATCCATCATGAGCTCTAAAGGGCCTAAAACCCCAACATCGTGCTGTGCGCGTCGCCCAACTGTGAACATGGGATTGGATGGATTGGAGGTGATTGAGACTCGAGCCAAGTTCCGAC
GCTGTACCTGGTGATGGAGTACGCCAAGTGCTGGACACTAACCAGGGAGTGCTTGGAAGAGTGTACCAGGGCCAACATCGTGCTATGAGGGTGAGGAAGAAATATATTGTACACAGGAACCTCTTGCAGGGCCAAGATTG
AGATTGTTTCGGCCTTCAGCAACGAGTTCAGCGTGATGGTATCGACAAGGGGCGCGTGAGAACCTCTTGCATGGCCGGACCTGAGAACCTCTTGCTGGATGCCCGAGCGTGTTTCAGGGCAAGAAGT
CTGACTTTGGCTTCAGCAACGAGTTCAGCGTGATGGTATCGACAAGGGGCGCGTGAGAACCTCTTGCATGGCCGGACCTGAGAACCTCTTGCTGGATGCCCGAGCGTGTTTCAGGGCAAGAAGT
ACGACGGGCGGAGGTGGACATCGAGCTGGAGCTGGCCTTCTACAGTGTCAACGACTGTGAGAGCGTGAGGAGTTGAAGCCATATCCGACAGGATCATCTGAAGCACTCTGTGCTGAACCAGTAAACGCTGT
ACTCTCGAGCAAATCATGAAAGACAAATCATGGGGTATGGCTACAACACGGGCAGAGGGTGGGCGCCCCAGGGTGGTCCTTGCGGGAAGAAATCAACAGCAAGCCCCGGGAAGCGCCCAGGGTGGTCCTTGCGGGAAGAAATCAACAGCAAGCCCCT
AGAGAATTGAGGTGGATGGTGGTATGGGCTACACGAGGGGTGAGGAGGTGGTGAAGGCACCAGCCCAGGGCAGCGATAGCGAGGGCGGCCATAGCGAGGAAGGCCGGAAGAGGCTGCCAGGGCGCCGCCGAGGAGGTGGGAG
GCTGGGCAGGAAGACTGAGGAGGTGGTATGGGCTACACGAGGGGTGAGGAGGTGGTGAAGGCACCAGCCCAGGGCAGCGATAGCGAGGGCGGCCATAGCGAGGAAGGCCGGAAGAGGCTGCCAGGGCGCCGCCGAGGAGGTGGGAG
AGCAAAGGCACCAGACCAGCCGAGGGGAGCACAAGGGAGCAGCGAGTTCCTTCCTTCCTTCCACCACCGCCAGGGGAGCTGAAGGAGGGGAAGGCGCCGGAAGGCCGGAAGGCCGGAAGGCCGGAAGGCGGGGGAGTGGGAG
GCACCCCAAAGCAGCCCGAGCGAGCCCCTCCAGCCCCATGGTCAGACCCCCAGGCCCGGAGGCGGGAGCTGAAGGAGGGGAAGGCGCCGGAAGGCCGGAAGGCCGGAAGGCCGGAAGGCGGGGGAGTGGGAG
TCGAGGGGCTGCCCCTCCAGCGGAAGCCCCTCCAGCCCCATGGTCAGACCCCCAGGCCCGGAGGCGGGAGCTGAAGGAGGGGAAGGCGCCGGAAGGCCGGAAGGCCGGAAGGCCGGAAGGCGGGGGAGTGGGAG
AACCTCCCTCCTAGCACCCCTAGCATGATGAACCCCGCAGAAACACTACGTTTGCACAGACGCTGCGGGGGGCTCACTGTTGCCAAATGGAAAGAAAACAG
CTCAGCGCACCCCAGGGTGCCCCTGCCTCCAGTCAACACCTTCCCAGCTGCCTGGCACGGGGAGCGCCGGCGGAGCGAGCGGCGTCACTGTTGCCAAATGGAAAGAAAACAG
CACCTTCCATGGTGGCCCGGTTCCAGGTGCCAGGTCCGGGGCGGCCCCACCAACCTCTTCACCAAGGTGTCGACCTCCAAACTGACCGAGATTCCCTGCTTCCGGATTCCCTGCTCCAGAGAATC
CTGCACCCCGTTGCCCGGCGGGCCCCACCAACCTCTTCACCAAGGTGTCGACCTCCAAACTGACCGAGATTCCCTGCTTCCGGATTCCCTGCTCCAGAGAATC
GGGGGGGGACCTGAGGTCACAAGTTGCCCATCTACCTTGGGGATTCAAACGGAAAACCGGCTGTCGCCAGATTCCCTGCTTCCGGATTCCCTGCTGACCAGCGTCGCG
CCCTCCTGGAGGCCCCTGGAGGTCACAAGTTGCCCATCTACCTTGGGGATTCAAACGGAAAACCGGCTGTCGCCAGATTCCCTGCTTCCGGATTCCCTGCTGACCAGCGTCGCG

TGCGGGGCGGGCCCGAGCCCCTGTCCCACTTGCGAGGTCTGCGCCAGGCTGCCCGCGGCCAGGCTTGCTCTTCCGCGCCGTGTGCGGGCACCGC
CCTGGCCTTCCGCACCCTCGTCACCCGGGGGAAGGGGCCAGGGAGGGATGAGGGGCTCAGCGGGGGGAGGGCAGGGCAGGGCAGGGGAGTTCTCTTCGCGCCTTCCCTGTCCCTGTCCCGTTCTCT
CTACAGGAGGGGAAGGGGCCAGGGAGGGGATTCTCCCTTATCATCCCTTATCATCCCTTATCAGTTTCCCTGAATTATATTTGGGGCAAAGATTGTCCCCTGTGTTCTCT
GGGGCCGCTCAGCAGGAGGACTTTGGGGACCAGGGGCAGGGCAGGGGAGGGAAGAAGGATGAGGGGCTCAGCGGGGGGAGGGCAGGGCAGGGG
GGCACCTGAGGGAGCTGCTGAGCCTAAAGACTGAGGAAATCGGGAGTGGGGTCAGAGAGGCAGATTCGGGGGTCAAGACTGGGCAGGG
GCAGGGAGAGCTGCTGCCCCAGGCTGGCGGGGCCACTTGTACAAATCCTTGTACAAATCCTTGTACAAATCCTTGCAAATAAATTTAGGAGAGTCGCCCGGACTGCCGCTGTC
CCCCACTTCCTGCCCCAGGCTGGCGGGGCCACCCGACCCCTCCCTGTTGCCCAATAAATTTAGGAGAGTCCCCCCAATG
ACCTACGGTTTTAAGTTATTACACCCCGACCCCTGCCCCTCACCTGCAAATGAGTTAAAGAAGAGGCGTGGGAATCCAGCAGTTGGTTTTTCCTTCGGAGCCTCGGTTTTCTC
CTGACCCTAGGATTTTCCCTCCCTGCCCCTCACCTGCAAATGAGTTAAAGAAGAGGCGTGGGAATCCAGCAGTTGGTTTTTCCTTCGGAGCCTCGGTTTTCTC
ATCTGCAGAAATGGGAGCGGTTGGGGGTTGGGAAGTAAGGATGGTCGTGGAAGAAGGCAGGATGGAACTCGGCCTCATCCCGAGTCCTATATC
GGGCCCCCATTCATCCACTCACACTCCCAGCCACCATGTTACACTGGACTTCTTACTCCAGTAGTAAATTTATTCAATAAACAATCATTGAC
CCATG

>DCAMKL2_H_SEQID#NA_14
GCGGGCCAACTTTGCCTCTCCCCGGTCGGCTCCCGAGCGGGGACTTGTGAGGCGGTGGCCGGGTGGAGGCGGTCCCCGGCAGGCG
GGAGGCACGGAGCAAGTCGCACTGGACAATGACCTGGGCAGCTCGGCGGCTCGGGGCAGTCGGTCCCGGGCGCGGTCCGGGCTCCT
CGGCCCCAACCTGCGCGGAGAGGGGGATGCAGGCCAGGTGTCCGGGCGGGTTAAGGCCCCTCAGAGCTGGAGCAGTATGAGTCGGCACAAAAGGC
AGTCGGCCGGAACGTCTTTTGCGGAAGCGTCTTTTGCGGAAGGGGTCGCGGGAGAGGGCGCGGGAGAGGTCGCGGGAGAGGGCTCATCCCAGTCGGCGCACAGTGC
CGCGGCCGCGGGGTCGCGGGAGAGGGCCCACTGCAGCTTCTACGGCACGCGCAGCGCCTCAGCGCCCTCGTCTTCCGGCGCCTTTACGGACCGCTACTTCAAGG
GCCTGGTGTTGTTGCCCATCTGCAGCGACGCGCAGCGGTCACCGGTCACCGGAAGGTCACCGAAGGTGAGAGTTACGTGTGTGCCCTCATCTCTCGTGAAAGTGAAGTA
CCGCACTACTCACCACCATCGACGCGCAGCGGCGAAGGTCACCGGAAGGTCACCGAAGGTGAGAGTTACGTGTGTGCCCTCATCTCTCGTGAAAGTGAAGTA
AAGTCGATTCACACCACCAAAAATATTAATCAAACCCAAGTTAGTGACTGGAAGCCATTAAACTAGACCTCAAGATCTTCTGAATAAAAAGACTGC
AAGAAAGTAAAGATTTCATCAAACCCAAGTTAGTGACTGGAAGCCATTAAACTAGACCTCAAGATGGAAAGCAGGTTACTTG
TCATTCCTTTGAACAAGTCTTAACAGATCTTAACAGATCTTTATTGCATGTGACCAGAAAAATTTCGTTATGCCCAAGATGTCGATCATAGTGAATGTCGT
TCTTGCAAGACTTTTTGGTGATGACGATGTTTTATTGCATGTGACCAGAAAAATTTCGTTATGCCCAAGATGTCGATCATAGTGAATGTCGT
GTCCTGAAGTCATCTTATTCTGGATCCTCAGCAGTTATTCTGGATCCTAAATCTACGAAATCTCCAGTTCTCTGACCAGAAGTTCAGGAGATTAAGCAGATTTCTGTCTCATGG
CTCCCAGCAGCCAACTTTCTCACTCTTTGACCAGATCTCTGACCGTTGCATAAGTCTGAAGGTGTCATAAGTCGAAGTGTGAATGGAAACAGATGCTGAATCATCAACTCTCTTGA
CAGATCTCTTCAATGTAAACGGTGAAACGGTGAGCTTGAGCTTGACGGTGCATAAGTCAAAGGTGTTGCCCTAAAGGAGTTTGCCCTAAAGATTATAG
GAAATACAAAATTGGAAAGTGTGTTGTGGAAAGGAACACCTGATTGAAATGAAGTGTCAAATACTGCGCGAGTGAAAACATCCAATATCATTATGCTGGTGTCGAGGAGGATG
ACAAAGCCAAATGTTGTGGAAAGGAACACCTGATTGAAATGAAGTGTCAAATACTGCGCGAGTGAAAACATCCAATATCATTATGCTGGTGTCGAGGAGGATG
GAAACAGCAACTGAGCTCTTCTGGTGAATGCCCTCAGGTATCTCCATGCCCTCAGCTGTGTTAGAAGGCCCTTTATACACTCTTGGTGTGTGGCTCGGAGATATCCT
GCCATGGTGTACAACTTAGCCAAGTCGGCTTGGGGACTTTGGGGAGACTTTGGGGCTTGAAAGGTGGACATTGGGCTAGAAGGTGGCACATTGGGCTAGAAGGTGGACATTGGGCTAGAAGTCCAGAA
GATGGAACCAAGTCTTTGAAACTGGCTATGGGCTATGGGCTGTGTAGAAGGCCCTTTATACACTCTTGGTGTGTGGCTCGGAGAAC
ATCATTGCTGAAACTGGCTATGGCCTATGGGCTGTGTAGAAGGCCCTTTATACACTCTTGGTGTGTGGCTCCAAGGAAGTAATCAGT
AATCTCCAGGAAGAATCTCTTCGACCAGATCTCTTGGCTGGGTACCGCGGACAAATCCTGACTGAATAATGCCTAACATCAGGGTCTCCGTCATCAAGGAATAACACGGCTCTAGATA
CAAATGCTTCAGGTAAATGTTGAAGTGTTGAAGCTGGTTCCGGCGTGTCAGATGATGCCTCGGCGTCTCCGTCATCATGAACACGGCTCTAGATA
TGAGGTGACAGGTAAACTAAAACAGCACTTTAATAATGCGCTCCCAAAACACTAACAGAACACTACCACCGGGTCTCCAGGAGAATAACACGGCTCTAGATA
AGGAGGGGCAGATTTTCTGCAGCACTGTCAAGACAGCGGGCAGCCCATCTCCCCGATGGGAGCCCATCTCTCCCCAGTTCCTCCCCGTTCCTCCCCAGTTCCTCCCCGTTCCTCCCCAGTCCCCGTG

```
CCTGGGGAAGCAGTCCGGCGCCCCACCCCTCCACCCCCACTGTCCTCCCGCTGCCCGGTGGTGAGCGGGCAGGAACCTGGCGCCGCCAC
CGAGACTGAGCCTCCTGCAGACGGGCGAAGCGGCTGCTGCAGGAAGCCAGCCTCTGCCGCCCTCTCGCCGCTGCTGCAGGCCTCCCTCTC
TTCACCGCCTGCGCCTGAGTTCGCCGGGTCCTGGCGGTGCCGGAGCCTGGCGGTGCCGGAGCCTGGCCTGGTGCT
```

>PIM2_H_SEQID#NA_15
```
CTGGCTCCGATTGGCTGTGGTACAGCAGGTGCCGGTCCGGATTGATCAAAGCCAGGTGGCGGGCGTGTCGCGCAGGTGACTCTCCCGGAGGCGCGTG
CCCAGGCTCAGTAGCATTGGCGTGGCCGCGGCGGCGAATCTCAACGCTGCGCCTGCGGCGCCTTCGGGGCGCCTTCAGTTTCTGCTTTCCACCCTGGC
GCCCCCCCAGCCCTGCTCCCCGAGCCTCCCCGGGACCCCCACGCGCCCCCGGGACCCCCAGCAGCGCCGCCCAGGAGCCAAGGATCGGAAGCGGTTCGAGGCGGAGTATCGACTCGGCC
GACCAAGCCTCTACAGGGCGCTCCCGGGGGGGCTTTGCACCGTCTTCGCAGGACACCGCGTCAAGTGGCCATCAAAGTGATTCCCGGAATCGTGTCGTGG
CCCTCCTGGGTAAGGGGGCGCTTTGCACCGTCTTCGCAGGACACCGCGTCTATGGAAAGTGGGTGCAGGTGTGGGCCACCTGGTGATCCGGCCTGCTTG
ACTGGTTTGAGACACAGGAAGGCTTCATGCGTGTCAGTCAGGCTGGCAGTAGTGCAGCCATCCAGCACTGCCTTTGCCCGCCAGGATCTCTTTGACTATATCACAGAGAAGGGCCCACTGGGTGAAG
GCCAAGCACGGCGCTGCTCTTTGGCCAAACTCATTGGTTCGTGGTGCCCTGCCACTGTCCATGGGCATCTGGTGTGGGGACATTCCCTTTGAGAG
ACCTACGCCGTGGCTGTCGACACCAGTACCATTGATTTGGTTCGTGGTGCCCTGCCATCCCGTGTGGGGACATTCCCTTTGAGAG
GGACCAGAGATTCTGGAAGAGATCCTGCTGGACCCTGATGCAAACACCAGCCGAGGATGTTACCCTCAAACCCTCCAAAGGGAGGCCCTTGCCCTGGT
CCTTGCTACCCTAAGCCTGGCCTGGCCTGGCCCCCAATGGTCAGGGGCCCCCAATGGTCAGTAGATGGGACATTTGTTGACTTGG
TTTACAGGTCATTACCAGTCATTAAAGTCCCAGAACCTTTGCTTTGCTGTGTCTCCTGATTTGGAGGGGAACTTCTCATTTGCTCTCCACCACCACACTTACTTTTACTTGGG
CAATCCTACAAAGGAGCCTTCCTCCCGGGACTCTTATTGATGATGTGTCACCCACACTTTTACTTCATATGCTTTTACTTGGG
GTTCCCATTTTGAGCCCCGGGACTCTTATTTGATGATGTGTCACCCACACAAACTTAGTTCATATGCTCTTTACTTGGG
CAAGGGTGCTTTCCTCCAATACCCCAGTAGCTTTTATTTTAGTAAGGGACCCTTCCCTGGGAGGAGAATGGGACCCTACTTTGTTATCCC
AGCCCAGGGATTTTTATTTCGGTGAGAAGAACCTTAATTCCATAATTTGGGAAGAGGGCTAGGGGAAATAAGGCTTGCTGTTTCCTGGGGACACCAATAGGATGGATGGAT
GGTTTTTGGGGGATGGCTAGGGGATGGCTAGGGGAAATAAGGCTTGCTGTTTCCTGGGGACACCAATAGGATGGATGAT
TTGTCCAATTACTAAATGTAAATAATCACGTATTGTGGGGAGGGGAGTTCCAAGTGTGCCCTCCTTTTTTCCTGCCGATTATTTAAAAGCCATGTGT
GGAAACCCACTATTTAATAAAAGTAATAGAATCAGAA
```

>PIM3_H_SEQID#NA_16
```
GCGGACGACGCGACACGCCGTGCGCCTACCGGGCTGCGCTACGAAAACGAGTCCCGGAGCGGGCCCGGCCCTCGCCGCCCGGAA
GACAGGCGCCAAGCTGCCGCCGTCTCCCAGCTAGCGCCGGCTGCGGGCGGGCCCGGCGGAAGGGGGTCCGGATTGCGCCCGCGCC
CCCGCGGAGGGATACGCGCGGCCCAAAACCCCGGGCGGCGGGTGAGGCGCTCACGCGGGTCCCGCGCCTGG
GGCCTTCCGGGCCCACTGCGCCGGGACCGGCTCGGCGCCCGGGTGTCGCCCGGGCGCCGGGCGTGGACCACCTCCCGGTGA
GGCTCGGGGGCTCCGGGGAGGCCGGGATGCGTCGCTCTCCAAGTTCGACGGGCGGCGTGGTGGTAGCGCGGG
AGATCCTGCAGCCTGCAGCATCGCCGACGGCGGACGACGTGGTGAAGCGCGTCATCCGGCCCCTGGAG
TAGCCGCGATCGCCGAGCGGGCTCCCGGTGGCGGCGCAGCCGAGTGGGCAGCCGAGTGGGCAGCCGAGCGCCTGGTCTCTGTGCGCTGGA
GTGGTGCTGCTGCGCAAGGTTGGGCGCGCAGCCTCTGGTTGGCGCGCCGCTTCTCGCAGGTGCTGCGCCCGT
GCGCGCCACTGCCCACCAGCTGCCGGGGTCGTGCGCTGCGGGGATCGCCGACACATTAAGGACGCAAAATCTGTTGTGGACCTGGGGACGCTCAAGCTCATCGAGACTTCGGTTC
```

```
GGGTGCGCTGCTCAAGGACACGGTCTACACCGACTTCGACGGGCACCGAGTGTACAGCCCCCGGAGTGGATCCGCTACCACCGCTACCACGGCGGCCTCGG
CCACCGTGTGTGGTCGCTGGGCGTGGGCGTGCTTCTCTAGATGGTGTGTGGGGAACATCCCTTGAGCAGGAGGACGAGGAGATCCTCCGAGCGCGCCTGCTCTTCCGGA
GGAGGGTCTCTCCAGAGTGGGGCGCCCCGAGAGCTGGCCCGGCCCTGATCAGATTGGGCGCCATCCCTGGATGC
TGGGGGCTGACGGGGCGCGCCCGAGAGCTGTGAACCTGCACCCTGATGACGTGGCCAGCGAGAGCTTGTGA
GGAGCTGCACCTGACCTCTGGTGGGGAGCTAGGGGACCACCTGCCTTGGCACGTCGTGTCTGTCCTCCTCGGG
AAGCAGTGACCTCTGACCCCTGGTGGGTGGGTGGATACTTGAGTGCCTTGAACGTGCTGTGTCGTCCAAAGACGCTC
CGGTCGAGGTCCCGCGGCCAGGGTCCCAGCCGAGCGGCCCCTCCCTCGCGTGCGGTGCGTCTGCGGCACGTCCTGCACACACAATGCAAGTCCTGGCCTCCGC
CTCTTTCCGCCGGCCCACGGGCCACGCAGCGTCCCCAGCCGCTACCCGGCGCCAAACTCTGTTATTTATTTATGTGGTGACCCCGAGGTGCCCTGCCCTCGCGTGCTTTAATTT
GCCCGCCCGACCCCAGGCCGAGCCGTACCGCCGCCCCAACTCTGTTCGCCTCTCCCAGCCCGTGTCAGAAGATGAACATGTAGTGCCTAAGGCACTCCAGGGACCCCAGGC
ATTGTTGAGGTTATTTCCTCGACCAGTCTGCCTCGGCCCGGCTGTCAGAAGATGAACATGTAGTGCCTAAGGCACTCGGGTGGGCCAGAGCCTGAGCGTTAATTATTCAG
CCTGATTCCGTGTCCTGGCCTGGCCCGGCCGTGTCTGGCCCGCCTGTCAGAAGTCTGCTTGTGGGCTTCAGTTCAAAGTGAGATGTCTGGAGATCATATTTTTTTATACAGGTATT
ACTGTGCCCAGGTTTGGGTTTAAATTATTGACTTGTACATAGTGAAAAAAAAAAAAAAAAAAAAAAACTCGAG
TACCTGTGTTTGTGTAGACATAGTGAAAAAAAAAAAAAAAAAAAAAAACTCGAG
TCAATTAAAATGTTTTTGTACATAGTGAAAAAAAAAAAAAAAAAAAAAAACTCGAG

>TSSK4_H_SEQID#NA_17
TTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGGATCTCGGCTCACTGCAAGCTCCGCCTCCTGGGTTCACGCCATTCTCCT
GCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCCACCACCACGCCCGGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCGTTTAGCCGGGAT
GGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCACTTCATTCTCAAGTTT
TGTTGGCCAACGATGGATAGAGAACATGGAGTCGTAGTCTCCAGTTCCTAGTGGAGGGCTGAGTCACATCCCAGAAGTAACAACAGCAGTGGA
GACAAAAAGAAACTGCTTCTCTTTCTTTCCCCTCCAGCTCAGCGAACCCTTCAGGTGTTGGCCTTTGGATAGGAGGCTTCAAGTAGTAAAGTCCCTGCTCTCAGCAAGCCAAC
ACCTACTTCACTTTGATACCCTCAGCCTTCAGCACCTTAGAGGCAGCAGCACCACCAAGCAGAAGGTTATGGTGGCCATTGGGCAAGGCCATTGGCCA
ACCATGGGAAGGGAGAGATGTCTGGTATATGAGGCTTTCTACCACAAAGCAGAAGGTTATGGTGGCCACAGTTGTGGCAGTGCTCTGATGACTATCTTAA
TGGCTCCTATGGGTCGGTATATGAGGCTTTCTACCACAAAGTCTTGCGGCACAAGTACCTCACTCAACTTCTATCGGCCATTCTCATCGGCCATTCTCATCAT
CAAGTTCCTGCCCCGTGAAATACAGTTAATGAAATACAGTTAATGTCCTTGAATGATCCAGCGGCACAGTGGTTCTCCCAGCTGACCCTGGG
TCTGGAACTGGCTCAGGGTGGTGATGTCCTTGAATGTCCTTGAACCGGGACTTAAAGTTGGAGAAACTGGTTCTGCTGACAAGTGGAATGTGAAGATATCAGACTTTGGCT
CATTGCTACCTGCACAGCAAGAGCATCGTGTTCTTCTAACCGGCCTTGTGTGGCCTGTGGTTGTCCTTCTCCTTCTTGGCTCAGCTTTGC
TTGCCAAGATGTGGCCTTCTAACGATCTTACGAGGCTTGCCTACAACCCTTTCCTGCTCATCTTTACACTCTAGTGGTCGCCATCTG
TTAACGCTGCCAGAGATCTTACGAGGCTTGCCTACAACCCTTTCCTGCTCATCTTTCAACATCTCCAGGAGTGCAAGAACCTGAT
CCCTTTGATGACACCATTTGATACCCTCAAAAAGTGCTAAGGCGTGCCACTAAGCGTGCCACCACACCAATCTGATACTTTAAACAGCACCAATCTCTAAAAATCTTTTTTTACCAAAATAAATCTGATTTTAGTTTCATCAAAAAA
CCTCCAGAATGCTACGCCAAGCCACTAAGCGTGCCACCACACCAATCTAAAAATCTTTTTTTACCAAAATAAATCTGATTTTAGTTTCATCAAAAAA
CAGGCTGCGTTGAGGCCATTGGCCAGCTCCACAACAGCACCAATCTAAAAATCTTTTTTTACCAAAATAAATCTAAGTCTGATTTTAGTTTCATCAAAAAA
GAGCAAAGCAGGAGAGGTCTTGGCCTAAAAATCTTTTTTTACCAAAATAAATCTAAGTCTGATTTTAGTTTCATCAAAAAA

>CKIL2_H_SEQID#NA_18
GCGGGCGCGGCGCAGGAGGGGAGCGGGGAGCAGGTGCTGGCCACAAGAGCAGCCGGCAGCAGTAACAGCAGCGGCCGCCGCCGCCG
CCGCCCAGTAAACGCGGACCGTACCCCAGGGGACTACCCGCCGCCGCCGGTCGGGCGGCGGGGGGATGGGCAG
GCAGTGGCGGTCCGGTCCGGCTGCCGAGGGTTAACCCCGGGCCTCCGGTCTCCGGTCGTCCGACCGGCCCAGGCCCCTCAGAGCCCTCTGCTGGACCGGCGATATCCGA
GTTAAATGGAAACCACCCTCTGGGAACTGGACTGGATGCCTGTTGTAGCTGTTCTACCATATCAGTGATTGCCATATCAGTGTATTGCAATGCAAGTGGGGGAGGGATGGGGGGAGGATCCTCGA
```

```
CTCATTTTTAAGCTGCCTTTGAAACAAAAGAGTTGAGGATAGGAGAGATAGAATGAATTTTAGGGGGGTTGCCTAATTTTTTTAAGCCTCAATTCAAAGATT
ATATAGCAAAAGTGAAACTTCTTGTTTGATATTTCATTCAAAACTTTCCACCCTGAAGAGTCATTGATCAGATATTAGATTATATAGAAGTCTGTTTGCCA
GGGAGCCAGTATTCATGTATTTGCTTGTGTGTTTATTTCGTGTTGTGTTTATTTGCTGTACTATCAAATCCAATTACTTTTTTTAGGTCCCCCTGA
TTTTTTTAAAATTTGTATGTCCGTCTGATTCAATTACATTTTTATTATGAAAAATAAGAAAGTTAGGTAAGAAGAAAGTCAACTGAGCTATTTTGCA
TTTTTTTTTAGAGCAAGAGTTCTTAACATATTACATTTTATTATGTGGTTAATAGCCTTCTGGGTTCCAACTGTTTGTAATATCAGACTGTTTGTAAGTTGAGT
GGCTTTAGTGTGTTTAGGGAGGAAAGAAAGAAGTGTGGTTAATAGCCTTCTGGGTTCCAAGTCTTGTAAGTGTCCTTAGAAACTTGAGT
AGAAGGAAAACTGACCAGAGTAGAACTGCTTCAGTTCTCTTAAGTCTCTAAGTATATTAGTGTCTTAAGTCTCTAATATCAGA
GGACTGTGATTTGAGATTTGAGAATTAGAAATAAATTTCTTTTTATTGTTTATTACCTGCTCTCTGTGTTTATTACCTGTGTTTATT
GGGAAATAAATTTAGGCAAATATTTCTTTTTATTACCTGCTCTCTGTGTTTATTACCTGCTTCTTGGCTTCTTGGCCTGATTTTCATACAATCT
CAATTTACGAAGCTGTAAAGAGGAAGAGATATTTGTTTCTAAATCTCAACTCTTCTAATAGGAAGTCATCAGGCAAATGAAAGTCTACCAGACTTTAAAATGGGCTGTTT
TTATACTTCTCTAGGTGTTTTGTGTTGTTGTGTTGCCAAACCTAATACAGTGGAAACCTTATTAAGGTCAAATTTGGTCTTTCTG
TTTGACCTTGTGTTTGTTGCTGCCCTGAAGCTGCACTGAACTTCTCTTTATCTGTGTTCTTTATCTGTTCAAACATATGTGCTTTCTG
CATCAAAGCTGCCCTGAAGCTGCACTGAAGTGTAGCATAGTGTGGAAATCTAAAAAAAAACTCAAAACACTATTGAGGCATATAACGT
CTCTTATAAGAAAATGTAGCAATAGTTGTGGAAATCTTAATTCTCTCCAGTTCAAACACTCAAACACTCAAACATTGCTGTGTTTATCTAAAG
CAACTGTAATATTGGAAGATCAGTCCTTCTGTATATTGTATAATTGTGTATAGTTGCTATAACACTACTTGCATGTCTTCATGTAATTATAATATTTATAAAT
ATATAGAGAGACATATCCTTATATAGAGTCATT


>PCTAIRE3_H_SEQD#NA_19
CTTCCTCCTCCTCCTCCTCCTCCTCCTCCCCACCTTCCTCCCTCCCTCCGGCGCCAGCTCAGGTTGCAG
CTTCTCTGGGGAACTGCTCACCTTTCCGGAGGAAGCTGCCCGTGCCCGGGAGGGAGGAGAGCTGCGTGAGTCC
CGCGGGGCGGTGCGCCTGGGAGGAGGGGGAGGTCGGAGGCCGGTTCTCCCGTGCAGTGCCCGGGTATAAGGAGGACACGCGCTGTGAGTCC
CATGAACAAGATCAGAGATGAAGACTTTAAGCCGGAATCCTTGGCTGAATTCACGGAGCAATTCAACCA
GCTCCACAACCGGCGAATGAGAACTTGCAGCTCGGTCCTCTTGGCAGAGTGCAGCACCCCGCGAGAGCAGCGGGGAGGAGC
CGGGGCAGCTCCCCTGGCGTCCAGTTCCAGCGGGCAGAACCAGCGCGCCTTCTCCATGGAGGCCTCTGCGCCCCGGCAGGTGGCA
GGATGCACGCACAAGGGTGTGCACAGGAGGCAGCTACAGCCCAGATCTGCCCAGGGCTCTCTCTGCCCATGGATATCCGCCTGCCCCA
GGAATTCCTACAGAAGCTACAGATGGGAGAGAGCCCCAGGAGCCACCTATGCCAAACTGACCAGTCTTCAAAGGGCGCAGCAAACTGAGCAGAAACT
GGAAACATACGTGGAAACTGGACAAACTGGGAGAGGAGCGCCCTGCACCTGGTGTCTCGCTGAAGCAGTATCTGGACCTGGGAAACCTGCATGCACAACGT
CAAGATTTTCATGTTCCAGCTGCGCCGACTTGGCCGGGCCTCGGGGCCTCGCGCCAGGGCCAGGGCAGGGCCCCTTCTCTGAGTTCGGCATGCACAAGAG
GGGGGAGCTGAAGCTGCACCTGCGGCGACTTTGGACTGGCCAGGGCCAGGTGGGCCACCCCACAGCGTGGCATCCAAGAGACGTGGCATCCAAGAGACGTGGCATC
ATGTGTCTGTGGATCCACAGGAGTACTCCACCCATCTTTCGCCTCCACCCCATTGATATGTGGGGACCCCACAGAAGAGACGTGGCATCCAAGAGAC
CCACAGTCAAGGAGGAGCTGCACCCTCACCTCATCTTTGCCTCCTCACACGGCGCCTCACTCTCCTGCTCCTCCATCTTCTCCTGAAGG
GCTCGCAAGTCAAGGAGGAAGGACCTCCCCGCAGCTGCCCGCAGTCTACCACGCGCCTCACCTCAGGGCGCCAGGCCTTCCCGGGCT
GTCCCCCTGCTACCCTGCCCGGAGGACCTGCTGCCCTGAGTCACGCGCCTACTGTCCTCCATCTTCTCCTGAAGG
AGATCCAGTCAGGAGGACAAGAGACCCAGGCTACCGAGGCTTGGCGGTGTCTTTGTGCCTGTTTCCCGTGTGTCCTCCCAGTTCAATCCAAGA
CTTGCTGTGGCCAAGGACCGCTTGGCACGGCTGCCCGGGGCGCAGGCAGGGAAGAACAGGGCGGAGCTTGCCCAGTTCCCCCAGTAGCCCACGCCAC
AGCCTGGAAGACCGGTTGGGGCTGGCGGAGAGCTGGAGGGGGTGGCTTTGGCTGTTGTTCTTTGTGTCCTCCCAGTAGCCCTCCTTT
ACCCACGTTGGGCGTGGCGGCATAAGCTGCTTCCCGGAGGACATGAGGGGGGGGTTGGGCGTCCTCGTACCCTCGTGTTTGGGCACCTGGGGAT
```

GCACACGGATGACAGAATCAAGGCGGCCAGGATGGCCACTCTGCCCTGGATACAGGCTCTACCCTCCTCCCCAGGAGACCTGCCTAGTGCCAGTTTGGTAGTCC
CCCTTTCTCGGCCCCTTGGAGCCACACACGTTCATCTTTTCCCCTCTGAGAGCAAGAAGAGACATGGCATGTTCTCTGGGACCCTGGAATCCTAGGTACCC
ACATGTGTGCCAAAGCCTACCCCACCTGGCCAGGTGTCCACAGCAACAGCAAGGAATAGTAGTCCCACTCTTTCATCAGCCTCTTTCCAGAGTGGCCTGGGACGGCCAGTT
GACACCCCTCTGGCTTGAACCATGGCTGAGCAGTGCCGGCATACGCTTTGCCGGCCATGCTGGATGCCAGCTGTGTCCAGAGGTGGCTACTGCCCTTGGAGCTGTGA
GCACGCCTGCCACCTCAGCCAGCCCCGCCAGCTCATCAGTGCTTACGGGCTCCTGATTAGCCAGGCTCTGATTAGCCAGCATCCTGTCCTACCTTCCCAGCCCTTGTGCTCGTCC
CCGGCTCCTGCCTGTCCGTCCACCCTTCCGAGGTGGCCGGGCTGCCGGTGCCCAGGCTGGTTGCCGGGGACCTGGTCTCGGGGACGGCACCC
AGATATGCAGAGTCACCTGACACTGGTGCCGGGCTGACCTCAGCTCAGGTGGCTGAACTCAGCTCCGAAGGCTCGCACAGCAGGGGAGGTGTAGATGCCACCATCTGAGTGGCTCTGTCC
ACCTGATCCCAATACCAGCTTCCCCAGCGTTCCCCCAGGGGCGGCCACGACAGGGAGGAGGTGTAGATGCCACCATCTGAGGGAGGAACGTGGA
ACAGGAGCAGGCTCTGATGCTGAATGGCACAGGCCAAGGGGAGCCAGCCTGCCTTCTACCCAGGAGGCTGTCCCTCGTATCCCCAGGTCTC
CTCAACACTGGGCTGATCCTGAACGTTCGTGTGACTCGTGGTCAGTGCAGAGACAGATAATATATTTAATTCATGTACAGAAA

>PFTAIRE2_H_SEQID#NA_20
GGGAAAAGGTTCAAGTGCGGGGTTTTCTCCTTGAACCTACAAGATTATGGGTCAAAGACTGTGTGCAAAGAGCTGGATGCAGCTGTACAGCCTGCTACCCATTGTT
CAGAGGGAGGCGAGGCACACAGCTGTCGGAGGCAGTCAGCCTGAGGAGGTCAACAGACCTCAAGCTAAAAGAAGCATCATGTTCCATGAC
TTCAATTTCACCCCAGGGACTTCAAGCTGCCCGTGCCCGTGCAGCCTCATCTCTTACTGAATTGAGAGCTGGATCTGAGGC
AGGGTTTTCAGTGGAGGAAGAGCCTCCCTTTGGGGCAGCCTCATCTTACTGATTGTTTTCAGGAAGAGGATCTGAGGC
TTAGCAGAATAAATGACAACTAGTGCTTAAAGTCATCAGCAGTGAAGAGTCCCATTTACAGCAGTTTACAAGGGGA
GGTTTTGAAACATGCCAATATTGTCTGCTCCTGCATGACATAATCCACACAAAGACACTGACAGTTTTTGAATACATGTCGTGGGGGCCTGTGCCGCGTACTATCACAGAGCTCTCCTGAAG
ATGTCTCAGCATCCAGGAGGCTTCATCCTCATAATGTCACCTGGTACAGTTTTCAACTTTTGGTCTTGCCGCGGCTACATCACAGACTGGCCAGTAT
GGGACCTGAAACCTCAGAACCTTGACCCCTGCTTGCTGGGGAGGGCCCTGATGCTTTGCCAACTATCCTGAACAGCTGGAGAAAAATCTGGAGACCTGGATACTT
CTTCAGAAGTCGTTGACCCTGCTCCAGCTCCAGCCTTTGTTCCTGGGGTTCCAACATCCTGAACAGCTGGAGTCCTACAGAGGATACTTG
TGAAATGTTCCAGGTGTCCAGGTCTCCAAGCTAAAAGGCTTTCCACACCCAGGAGACCCGGTCTAAAGCCAGAAAGTCACCA
GCCGGGAGTCTCCAAGCTACCTAACTACAATCCAGAATGCTAAAAGGCTTTCCTCAGTTTCATGATTATTTCAGCGGGCCCTGCCATCTCA
GCTGTACCCAGTTTAGCAGTTGTTACAGTTCGTGGTGAAAGAAAGGGCGAGATCTTTGTTACAGTTTCAGGGCTTTGCGTTTCATTTGCTTCAGCTTA
CCCAGCCCAGTTTAGCAGTTGTTACAGTTCGTGGTGAAAGAAAGGGCGAGATCTTTATGTCCTCACCTGGAATAATGGTGTTCAAGGCAATAGTACATAA
CTAAGAAGCTTCAAATCTAACTCCATACTGAACAAGGGGCTTTATGTCCTCACCTGGAATAATGGTGTTCAAGGCAATAGTACATAA
TAGTGGAAGAAAATTCAGTGAAGGTTATTGTCATTGCATGAAGTTGATTGATTAAAAAAACTGACATAAAACTAAGTCTAAGAAGAGAGG
GCTTCTGAATAACTTCATTTGCTTACAGAGACCTGCCACTCCTGTGTTCACTTACAATTTCTCCCAACAAAAAGGATTTCTGTGGCAAGGAGAAAAACTAGCCTAA
TCTGGAGACAGAGCTGGTGTGCTTACAGAGACCTGCCACTCCTGTGTTCACTTACAATTTCTCCCAACAAAAAGGATTTCTGTGGCAAGGAGAAAAACTAGCCTAA
AGTTTAGAAAAAGATGGTTTCTTGGATAACATTGTTAACAAAAGGACTTCCTGAGGACTATTTATTATTTGTAGGGGGCACTTCCTCCTTACTAAACT
GGATTATCAGAACGTATTTACTGTTAACAAAGGACTTCCTGAGGACTATTTATTATTTGTAGGGGGCACTTCCTCCTTACTAAACT
CGTATTACCACCCAACCCATATTTTTAGAAGTTTGGCTGCTGATTTAGAAATGAAATTCAGGCTATCTCCTCTACTGAAATCTTAATTGTCTGTTGATGAT
GGCAACAGTGATGGGGAATCTAAGCGCTGTTTGTTTGATGGTGGAGATATTGAATAAGTAGTTGGGAGCAGAGTTCCAAAGATGTAATTTAAGCCCTGTTGCAC
CACC

>ERK7_H_SEQID#NA_21

GCGGGCGGCTCCTCGGCCGCCATGTGCACCGTAGTGGACCCTCGCATTGTCCGGAGATACCTACTCAGGCGGCGCAGCTCCGGGCAGGGGGCCTATGGCATTGTGTG
GAAGGCAGTGGACCGGAGGACTGGTGAGGTCGTTGGCCATCAAGAAATCTTTGATGCTTTAGGGATAAGACAGATGCCCAGGACATGGCTTCCTTCTTG
CTCCACCACCACACACCTGTGTTTCTGTCTCTTCAGAGAAACATTCCGGGAAATCCGGGATCATCACGCTCTTCCTCCAGGAGTTTGGGGACCATCCCAACATCAGCCTCCT
TGAACGTGATCCGGGCAGAGAACGACAGGGACCATTTACCTGGTGTTTGAGTTTATGGACACTGACCTGAACGCAGTCATCCGGGGACCAGAAGCCGTCCAATGTGCTCC
ACGTCCACGTGCGCCTCCATCTTCTACCAGCTCCTGCGGGCCACCGGTTCCTCCACTCGGGGGACCTCCTGAGGACCAGGCGGTTGTGCACCGGAGACCGTCCAATGTGCTG
TGGATGCCAACTGCACAGTGAAGCTGTGTGACTTGGCCTGGCCTCCTGGCCGGCCGATACACCCTTGGGGTGGACATGTGGACTGTGGGTGTATCTGGGGGAGATGCTG
GCCACACGCTGGTACCGGACCACCGGAGCTGCTGCTCTCCTCCACCAGCTGGAGCTGGAGCATCTGAGACCGCCATCTGATGGAGGAACCTCCTGGCGCTCTCGCGC
CGGGGGAGACCCCGTTCCCGGCACGTCCCGCTGTCCTGTGCCACCAGCTGGGGTCCCGGACACGTGGGCCCGGGACAAGCGGTTAAGGCGACAAGAGAGGGGTCCAGCAGTG
TCAGGCTGCCGTGCCTGTCTGCGCACCAGCTGGGTCCGCCACCAGCTGGGTCCCGGGACAAGCGGTTAAGGCGACACGAGAGGGGTCCAGCAGTG
CTTAGGCGACTCCTGGTGTTGTTGCCCCCGGACTACTGCCCCGGGAGCCCGGGACAAGCGGTTAAGGCGACACGAGAGTGTG
GGCACGAGAGGCAGCAGATGTGCGGCCCGGGACAAGAGGGGCCCGGGAGAGGGTGTCTCCCAAGCCAGGTCTCTGTGCCTGAGTACCGCGCAGCCCTCTTCTCAGGCAGAACT
ACCTGTGCAGGGTCCCAGAGTCCAAACTGTCTCCTGACCTCCCGGCTTGCGCCCTGACCTCCCGGCTCCCCCTGCTCCCCCTTGACCGGGGCTCTCGGTTGGTGAAGCCAAGCGGG
CGGCTCCCCTGTCGCCCTGACCTCCCGGCTTCCTCCCGGCTCCCCCTGCCCCTCCCGGCTGGCCAACCAGGCCGTGGGGTGGGTGAGGGTGGC
AGGGGGAGCTGCGCCTCCGTGACCTGCGCGCTCCAGCTCCCGGCCGCCGCCGGGGTGCCCTCTGCCCTTCGCAGGGGTCACCATGTGTGAGGCGCCCTACTCC
CTTCACCTGGCCCCTCGTTCTGCCCCAGCCCCTTCCCCAGACCCCTCTCCCCTGCTCCTGCCCCTGTTCTCTGCTCCCCTTAGCCCTGCCACCCGTGTGCCTCGGCCCGTTGAAGTTCATC
GGGAGCTTGCCCGGGGTCTCCTCGGGGGAGCAGATGAGGGCCGTGCTCTCCTCGGGGGAGCAGATGAGGGCCCTGCCCCCACTGACTTCCTCCCAATAAAGTCATC

>CKIIα-rs_H_SEQID#NA_22
AGGGGAGGCGCCACAGCCGCGCTGCCGCTTCCACCACAGTGTGAAGAAACAGGTCTTACCCCAAGCTGCTTCTGAACACAGTGACTG
CCAGATCTCCAAACATCAAGTCGGTTGGATCCAGTTTGTCCGCCAACCTGTCTGACATGTCCGGACCCGTGCCAAGCAGCAGTTTACACAGATGTTAATACACAC
AGACCTCGAGAATACTGGGATTACGAGTCAGTTGTGGAAATGGGAAATCAAGATGACTACCAGCTGGTTGGAAAATTAGGCCGAGGTAAATACACAGTGA
AGTATTTGAAGCCATCAACATCAACAATAAGTTGTGTTAAATTCTCAAGCACAGTTAAAAAGAGAAATTAAGCGTGAAATAAAGATTTTGG
AGAAATTGAGAGGAGGAGGTCCAACATCACTCACACAGACGTTCAACATTGTAAAAAGACCCGCTTGTGTTTTGAACACGTAAACAACACA
GACTTCAAGCAATTGTACCGACGTTCACGAGTTCATGGATTATTCACGAGCATGGGAATTATG
CACAGAGATGTCCAAGCCCCCATAAATGTCATGAAGTCACAGAGAAGCTAGCTGTGTGACTGGGGTTGGCTGAGTTTATCATCCTGGCCAAGAA
TATAAATGTCCGAGTTGCTTCCGGAAGGAGCCCAATTTTCCATGGAAGGAGTCACGTTCAATGATATCTTGGGCAGACACTGTGAAAGGTTGTATGC
TGGCAAGTATGATCTTTCGGAAGGAGCCCATTTTCCATGGAAGGAGTCACGTTCAATGATATCTTGGGCAGAGAAGATTTATATG
GCTATATTGACAAATACAACATTGAATAGATCCACGTTCAATTAGATCACGTGCGATATACTGCGATATGACAAACTGCTAGCAGCCCTATT
AGCACCCTGTCAGCCCTGAGGCCTTGGATTTCCTGGACAAACTGCTCGAAATGGGTTCATCTGGCAGGCTCAAAGGGGGCTCACCAGTACACCCGTTCTTCAG
TGCCAACCCCTTCACCCGTTGAAGGACCCAGCCTGTGCTGCCAAACCCCCTTGGATGCCGTCTGCTCGCGCTCAGTCAGTA
ACGGCCCTATCGTGTCTCGATGCGGGAGGTGGGGGAGTCCACCCTCCTGTGATGCAGTTGCGGCCGCGGGGGGTGAAACACTTCAGAAGC
AACGGTGTCGAACCGGTGCTTGTGTGGATTTATAGTGGATTTATAGTTCAGTCATAAAAATTATAATAGGCTGA

>DYRK4_H_SEQID#NA_23

ATGAACACCCCTCTGCAGTCTGCACATACACACCTTACCCTGGTCGAGTTCATAGCTACTTGGTGTCTGACTCCCCTTCCTGCTTCAACCTTGTCTCT
GCCCACGTGCAACTTTGATCCTGCCCCACTAAACTGGTGCTGGCACAATGCCTCCCGCAGCGGCGGCGGCGGCCTCAGCGCGGTCAGCGCTCCTCCGC
CGCCTATCCGCACCGGAACAAAGTAAGGCCGGGAGGCTCGTACTTCACGAGCAGTCAGGCGACGCAGGCGAGCGGCGGCGTGGTGCGCGACAAGAAATACA
GGTCGTGGGGGAGAATGCAAAAGCGTGCAGAATCGGAATGGCATCCCGGAATCGTGCTTCGTGATATTGCCAGAGAAATACAAAAATCTGATCTTC
GTTGTAGTTTTTGCAGCAGCTGCATCCCGGAAATCGTGCTTCGTGATATTGCCAGAGAAATACAAAAATCTGATCTTC
AGCCCAGGACTCAAATGGATGCTAAAAAGCCAAGGAAATGTGATTTGACTCGTTGGAGGATCATGCTGAGGGAAATCAGCAGAAGGGCAAG
CACCTCTGGGTACAGCGCAAAGGCAAAGGTGTGGGGAGGATCATGCTGAGGGAAATCAGCAGAAGGGCAAG
GCCTTAATCTAACAGGTGCATTAACAGAAGTCAGAAAACAAAGTATCTCCCAGGCCCTGAATGCATTAGAAATTGACTCCAGTCATCTGACC
CCCAGGCTCTTGCCCTTCCTCAAGGATGCGAGGCCTCCTTCCACCTTACTCCCTTGTGGACACCAAGGGGAAGAAGAATACGG
CACCTTCCAATATCAAGAACTCCAGAATGCTTCATAAGAACCAGTTACTTCACTCCCTTGTGGACACCAAGGGGAAGAAGAATACGG
TAAGCTTCCCACACATTAGCAAGAAGTCCTGGGAAAGCACAGTAGCTGTATCAGGTGAGTGCAGAACGGACTGGAGTGCAGGTGAAAGACCT
GGTTTAGTCCAAGGCTTCAGAAATACCTTCCACCTAGCATTAAAACCCAGGAAAGCACAGTAGCTGTATCAGGTGAGTGCAGAACGGACTGGA
AGAGCTCAAGGCTTCAGAAATACCTTCCACCTAGCATTAAAACCCAGGAAAGCACAGTAGCTGTATCAGGTGAGTGCAGAACGGAC
GCGGCAGAGCCCTAAGCTTTTAAGAACCAGTTGTCTCCATATGAACAAAGTGAAATCTGGGCTACGCGGAGCTGTGGTCTTGAAGCCAA
GAAGCTCGACACGGCTCCTGAGAAATTTAGCAAGACGAGTTTGATGATGGCCAAGTGCTTGAACATCACAAAAATCATCAGGAACAAG
AGTTCTGGAGACAATCGGAAGGGGTCCTTTGGACAGGTGGCCAAGTGCTTGAACATCACAAAAATCATCAGGAACAAG
AAGAGGTTCACCAGCAGCGGCCCTGATGGAGCTGAAGATCCTTGAGCTCCTGGAATCAACTTTCAGGCTTCAGTCTGTCCAT
TTTTCTACTTTCGGCCTTCACTCTCTCTGTTTTGAAGTGCTTGCAGATGCTTTGGGTAGAACACCAGAAAATAGTGCTATACCAA
AGTTCGGCGCTTCACTCTCTCTGTTTTGAAGTGCTTGCAGATGCTTTGGGTAGAACACCAGAAAATAGTGCTATACCAA
AAGGGCCAAGCCTTCGGGCCACCCTACGACGTGGCCATTGACGGTGGAGCCTGGCGCTACCCCTGTTCCCGGGGAGAAT
GTGAATCCTGGGCCACCCTACGACGTGGCCATTGACGGTGGAGCCTGGCGCTACCCCTGTTCCCGGGGAGAAT
GAGGTGGAGCAGCTGGCTGCATCATGGAGGTGCTGGGTCTGCCGCCAGCCGGCTTCATTCAGGACGACCTCCAGGAGACAGACATTCTTTGATTCCAAAGGT
TTTCCTAAAAATATAACAACAACAGGGGAAAAAGATACCCAGAATTCCAAGGACCCTCAAGGTGCTGAAAACCTATGACACCAGCTTCCTGGACTT
TCTCAGAAGGTGTTTGGTATGGGAACCTTCTCTTCGCATGACCCCCTCTGAGACAAGGAAGGACAAGGTTCAAGGCTGTCATCACTGAGCAGAAAGATGAGATCA
CAGGCCCCAGACCCGGAGGAAATCCAATTCCTTTTCCCCTCTGAGACAAGGAAGGACAAGGTTCAAGGCTGTCATCACTGAGCAGAAAGATGAGATCA
CCAAAGAGACTACAGAGAAAAACAAAAGATAGCCCCACGAAGAAGCATGTTCAGGTGATCAGCAGGCTGACACTGTTCAG
CTGCCTCAACTGGTAGACGCTCCCAAGAAGTCAGAGGCAGCTGTCCATGAGGTGTCCATGAGGGCGGAGTGTCCATGAGCAGAGCAAAAACTTCTCCCTCAAGAA
CACAAACGTTTACCCCTATTGTATGACCCTTTGCTGAGGGTATGTGCTGTCCTTGCTGCTTTGGGTAGTGTGGAGGGTATGTGCTGTCCTTGCTTGCTGAGGGTATGTGCTGTCCTTTCCACCAGTGATTTGTATTGTACAATACTTC
AGACTGTTTTTTTAAATACATAAAAACTTTATGTTAAAAACTCT

>HIPK1_H_SEQID#NA_24
CCGGCGGTAGCCCCGGACGGCAGCAGGAGGCCGAGGGGCGGGAGCGGGCGGGGCTGAGGGGCGCGGGAGGGGGGCTGGGTGGTGGAGCGGG
AGGGAGGCTGAGGAGGCTCCCGGGGACGGCTCCCGGGGGCCCGGGGGGCCGGGGCGCTCCGCGGGGGGGGTGGGGGGTATGGATGACC
CGGCTGCGGGGCCCCAGATCTCGCCCCATCAGTGCTCTCCGGACGTGCCTTCTGCAGTGCGAGTGCGGGTATGGCATCACAGCTGC
AAGTGTTTTCGCCCCATCAGTGCTCTCTGCAGTGCGAGTGCGGGGGATGTTTCAGGACAGAGTAGCAACG
ACAAAATATTATACCCACAGACAAAACCCTCCAGCCACAGAGGGCAAGCCAACTCTCTCACCAGGTAGCAACTTCAACATCCTGTTACGGACCAGGGC
CTCCTCCTCCCGAGCTCCTGCAGTGAGCATATTGTTGTAACAGCCGCTGATAGCTCGGGGTAGGGAGCTCCTGCAGTGAGCATATTGTTGTAACAGCCGCTGATAGCTCGGGG
CACAGAAGCAACGTTTCTTTGCTTGAGCCAATCAAAAATGTGAAAAAAGTTGACAGCAAGTGTCAGATCCATAGAAGA
ACATCCCCCTCTGCAAAAACAGGACTGGTGGGGTCGTGCTGCCACAACCACCACTGTGACCACCAAAGGTAGCAGTTCCAGGGGATTAGCAGAGAGAGAGAGGGGGATT

ACCAGCTGGTCCAGCATGAGGATCCTTTGCTCTATGACCAATAGCTATGAAGTCTTGGACAGGTGTGGCTAAGTGCTGGA
AGAGGAGCACCAAGGAAATTGTGGCTATTAAAATCTTGAAGAACCACCCTCCTATGCCAGGACAAGGACGATTGAAGTGAGCATCCTTTCCGCCTAAGC
AGTGAAAATGCTGATGAGTATAATTGGTCCGTTCATACGAGTGCTTTCAGCATAAGAATCACACCTGCCTTGTTTTGTTTTGAAATGTTGGAGCAGAACTTATATG
ATTTTCTAAAGCAAAACAAATTTAGCCCACTGCCATCAAGTACATCAGTTCGTTGATCACAGACCAATCTTGCAGCAGTGGCCACAGCCTGATGAAGCTCAAGAGTCTTGGTC
TGATCCAACGCTGACCTTAAGCCTGAAAACATCACGTTACTCACGGTTACTACAGAGCTCCTGAAATTATTCTTGGGTTACCATTTTGTGAAGCTATTGATATGTGCACTGGGC
CAAAGCTGTGTGTCTCAACCTACTTACAGTCACCTTACAGAGCTCCTGAAATTATTCTTGGGTTACCATTTTGTGAAGCTATTGATATGTGCACTGGGC
TGTGTTGATAGCTGAGCTGTTCCTGGGATGGCCTCTTTATCCTGGTGCTTCAACAGAGATCCTAATTTGGGTACCCACTGTGAAGAACCATGAACCTG
ATCTTCTCAGTGCCGGAACAAAAACAACCAGTTTTTCAACAGATCCTAATTTGGGTACCCACTGTGAAGAACCATGAACCTG
GAGACTGGAATAAAAATCAAAAGAAGCTCGGAAGTCAATTTTAATTGCTTAGAGACATGGCTTCTCAGGTGAATATGTCTACAGAACTGAATGGGACAACAGACAT
GTTGGCAGAGAAGGCAGAACGGAAGAGAATACATTGGATTTCCACATAGCAATCATGTTAAGACATGGAATACTAACATGGAGAGATCTGCAAGCGGAGGTTCAATACAGTGCA
ATCAGTTTGTGACAATGACTCACCTTTTGGATTTCCACATAGCAATCATGTTAAGACATGGAATACTAACATGGAGAGATCTGCAAGCGGAGGTTCAATACAGTGCA
TGATACAGTGAGTCAGATCAAGATCCCTTCACTGAGCTTCTAGCAGCTGTTGCCCAAATACAAGCACAAATCTAACTGAGCTTCACTACTAAAACTACCAGTCAGCTTTGTACC
CAATCAGGCCAGTGTTCTCCAGTTCCTCGAGCAGCTGCTACTGCCAGCAGGGTGTTTCCTTGCAGCCTGAACCACCAGACTCAGAGGAACCTGAACCCAGACAT
CATCATCTGCTGCACCAGTTCCTGCGGTTTCAAACTGACTACAAGCAACAAAAGCATTCTGAGGATGCAACTCTCCACCCTCAAGTAGCCACCATCA
TTATAGTATGTCCACCGCGTTTCAAACTGACTACAAGCAACAAAAGCATTCTGAGGATGCAACTCTCCACCCTCAAGTAGCCACCATCA
CACCAGCTGCTCAGCCAGTTCAGCTGGGTGCCTGCGGCCGGCCAACAGTTGCCGGGGTAGCTCTACACAACTCTGTTGAACAGAACTGCCCCACAGCAATGATTCCAGGCGTGGGAGGTGAC
AGCAAATTCCTGCCTTCAACTGGCAACAGTTGCCTGGGTAGCTCTACACAACTCTGTTGAACAGAACTGCCCCACAGCAATGATTCCAGGCGTGGGAGGTGAC
AGCAGTAGCTGACTGAGTTGGTGCCCATGTTGTCAGACAACAACCATCCAGTTCCTCCGGAAGAATAAGCAGTCAGCTCCAGTCTCTTCCAGTC
CTCCTCTGAATGTTCTGCCTTCCAAGTCTATTCTGGTTGGGAGCAGTCCCGTTGTTTATAATTCCTTGGTCCCGTCCAAGATCAGCATC
CTCTCTAGATGTTCTGCCTTCCAAGTCTATTCTGGTTGGGAGCAGTCCCGTTGTTTATAATTCCTTGGTCCCGTCCAAGATCAGCATC
AGCCCATCATCATTCCAGATACTCCCAGCCCTGAGTGTCATCACTGTCACTGTCAATGAGGAAGAGGACAACAAATACAAGCCCAGTAGC
TCTGGACTGAAGCCAAGGTCTAAATGTCATCAGTATGTCACTGTCACTGTCAATGAGGAAGAGGACAACAAATACAAGCCCAGTAGC
GTGCTCTCCGGAGGCAATAGTGGATCCGTTTGTGAGCCTGGCAGAGTTGTGCAGAAGTTAGAGGTGTGTCCTCTTTGAGCCTCACTGATACCCTGA
AGCTTGGTGACTGCACTGTAGCAACCAGGCTCTCTGAGCAATAAGACTAAGCCCAGTGAGTGCGCTTCAGTGAGGCAGTCATCGTGATGCTGTATCA
CCCCCACAGGGTATCGAGCTCAACGCGGGGGACCCAGTGCAGCAAAACCACTCAAATCTTAGCCAGAGTCCAACCTCACACTCGACAGGAG
AGAAGCAGCAACCCAGCCCCGCCCCGCCCCTGCTCACCTGCCAAGGCCAAGGTTCCTCAAGGCATGCTGCTGTCCGTGTTCCTGTCAGTGTTGGGCCCA
GCACCCCACACCTTGCCCGGGCCCCTGCTCACCTGCCAAGGCCAAGGTTCCTCAAGGCATGCTGCTGTCCGTGTTCCTGTCAGTGTTGGGCCCA
ATTGCTCATCTTTCTCCCCAGAGGTTCCTCTCAAGGCATGCTGCCTATACCACTCACCCTGGTCTTCCCGAGCTCAACAATTTACACTGATA
GCCTCCTCACTTCTGCCACCAGAGTCCTACCAGCCAGTATTCCTACTTATAGTGGTACAGTGCCCTGAAATTTCTTAGCCAGCAACGTTGGTGTCTTCCTGAGCATGGGGGAACC
CCGCTGAGTCCTACCAAGATCAGCCAGTATTCCTACTTATAGTGGTACAGTGCCCTGAAATTTCTTAGCCAGCAACGTTGGTGTCTTCCTGAGCATGGGGGAACC
ATTGGGCTATGGAGAGATCCTCATTGTTAGTCGTTGTTCTGTGAGTCAGTCCTATTTTTGTCTCTGAAATATTTTGTGTCTCTGATTTCTATAAATGCTTTAAAAACAA
TGTCTCAGTGTTGACTGCATTGTTAGTCGTTGTTCTGTGAGTCAGTCCTATTTTTGTCTCTGAAATATTTTGTGTCTCTGATTTCTATAAATGCTTTAAAAACAA
CCCATCTTCTGCCAGTTACCAAGGTTGTGGTCAGGAAGAGAGAAAATGCTGATAGAAGGAGTTGAAAATCGATGACATTTAATTGGACATTTAATTACTACAGAGTTTCAGTTTTG
GTGAAGCCCCTTTATTTCATTTATTATATACTTAATGGGCAATTGTTATTTTTGCAAAACTGTTTATGTGTTACTATTACGTATTACTCTTCATTGACATTTGGACATTCAGTAGTCTCGTGTTGTTGT
TTTTTGTTTGATTAAAAACTCAGAATTATACCTTAATTGGGCAATTGTTATTTTTGCAAAACTGTTTATGTGTTACTATTACGTATTACTCTTCATTGACATTTGGACATTCAGTAGTCTCGTGTTGTTGT
TATAAAGTAGTGTTTAAAAGGCAGCTCACCATTGCTGTAACTTAATGTGACGAATCCATATCGTGATTGAGAGAAATGCTGAAAACAACCAAGTATTCTTTTTAAATGAAGCA
CCATGAATCTTTTTTATTTTTTAAAAGTCTTTCTCTCTCGATTCAGCTTAAATTTTTATCGAAAAAGCCATTAAGGTTGGTTATTATTACATGGTGG

TGGTGGTTTATTATATGCAAATCTCTGTCTATTATGAGATACTGGCATTGATGAGCTTTGCCTAAAGATTAGTAGTATGAATTTCAGTAATAACACCTCTGTTTT
GCTCATCTCCCTGTTTATGTGATTTGTTGGGGAGAAAGCTAAAAAAACCAGATAAGAAACCTGAAACCAGATAAGAACATTCTTGTGTATAGCTTTTATACTTCAAA
GTAGCTTCCTTGTATGCCAGCAGCAAATTGAATGCTCTCTTATTAAGAGTCTATATAAGTGCATGTAGGAATTGCAAAAAATATTTAAAAATTTATTAC
TGAATTTAAAAATATTTAGAAGTTTGTAATGGTGGTGTTGTTTAAATATTTACATTGAATTGTGTATGTGACATTAGTGTGTATGTGAATCTTCAGTGTTATGTGTATCCGAGATGGATT
CTGCTAACCCAAAATGTTATTTGTAATCAAATGTGTAGTGTAGTACACTTGAATTGTGTATCTTAGTGTGTATGTGAATCTTCAGTGTTATCCGGAGATGGATT
GATGTCTCCATTGTATTTAAACCAAAATGAACTGATACTTGTTGGAAATGTATGTGAACTGATAATTATATTACTGTAGTGCTGAATGAGC
AGGGGCATTGCCTGCAAGGAGAGAGGAGACCCTTGGAATTGTTTTGCACAGGTGTGTTGGTGGTAGGAGTTTTCAGTTGTGTCTCTCTCCCTTTCCTCTCCTC
CTTCCCTTATTGTAGTGCCTTATGTAGTGCTTAATAGTAGTGGTTAATGAGTTTACAGTGACCTTAGGATGGACCAGCAAGCCCCGTGACCCTAAGTTGT
TCACCGGGATTTATCAGAACAGGATTAGTAGCTGTATTGTGTAATGTGTAAATGCATTGTCTCAGTTGTTCCCTGCCAACATTGAAAAATAAAAACAGCAGCTTTCTCCTT
TACCACCACCCTCTACCCCTTTCCATTTGGATTCTCGGCTGAGTTCTCACAGAAGCATTTCCCATGTGGCTCTCTCTCCCATGTGGCTCTCTCTCCCATGTGGCTTGCTACCTTGCTTCTGT
GAGAATTCAGGAAGCCAGTGAGAGGAGCCAGTGAGAGGAGACATAGTTGCACATAGTTGACACATAGTTGAAGAATATAGAGAAGG
ATGTGGCAGTCGTCCTTCCCATACTATGAGGAGAAGTTTGGTGCCGTGCCATCTTTTGAGTTTCTTGAGTTTCTTGTGTGAAGAATATAGAGAAGG
TGCTTACCATGTCCCATACTATGAGGAGAAGTTTGGTGCCGTTTCTTTTTAACAGCTTTGAGGTTTCTTGTGAATTGTCCTTGTACTCTCAGCT
AACCAAAGCCTGTTGAGTCGTCGGTCAGTAGAAGTTCCTGCCCATAAAGGATTTGGGGAAGAAGATTAATCCTAAATACAGGTGTGTT
CCTGCATGGATCTGGTCAGTAGAAGTTCCTGCCCATAAAGGATTTGGGGAAGAAGATTAATCCTAAATACAGGTGTGTT
CCATCYGAATTGAAAATGAATATATTTGAGATATAATTTTAGGACTGGTTCGTGTAGATAGGAGATGGTGCAGGATGGAGATGGGG
ATGGACCTGGTCAGTTGTCAGCCAGGTTGAACACCGAGGAGCTGTCAAAGTATTGGAAGTTCTTCATTGTAATTGAGTTGTTTCAGCCACTAATTGGTCAAGATAGCCAA
CAGGTAGTCCGTACAGCCTACCAGGAACATGTGTCTTCTTATTTTAAAATCATTATTGAGTTGGGTCTCCTTTTGCCAAGCACATTCTGATTTCTTGTTGGAA
GCAGTTGTATAATTTCTGTCACTAGTGTCATACAGTTTGTCAGTTTGTCCTTGTCTTTTTCTGTAAGGGACAAATTTTTTGTTCCTGTCTTTTTCTGTAAGGGACAAATTTTTTGTTAAGGGACAAAGAAAAC
CACAGGTCTAGTTTCTAAAGGACAAATTTTTTGTTCCTGTCTTTTTCTGTAAGGGACAAGATTTGTTAAGGGACAAAGAAAAC
CAAATCCCATTCCTGCACCCCAGTTCCAATAAGCAGATACCAGTTCAAAACTGCAAACTGCTAAACTCCACAGAAAAGGATAATACCAAGAGCTTGTATTGTTACCT
TAGTCACTTGCCTAGCAGTGTGTGTCCTCACTGTATGTCTCACTGTTAGATTCTGTCCCTGGGGTTAGCCTGCAAACTGCAAACCCGTGTKGGCCCTGACCAGGAAGAATCCACCTGTGT
CTGCTGAATGTGTATGTGTGCCTGGGTCACAGTGACCTGACCTGAGTGTGGCTTTAGATTCTGTCCCTGGGGTTAGCCTGCAAACTGCAAACCCGTGTKGGCCCTGACAGTTTGT
GAGCAGCTGGCCTGGGTCACAGTGACCTGACCTGGGTCACAGTGACCTTAAGTCCTCTCTCAGAACTGGCATTCCAACTTCTTCCTTCCGGGTGAGA
GAAGAGCGGAGAAGGGTTCAGTGTAGCCCACCTCAGTGTAGCCCACTCGGTCAAACAGTTTTAATAGCTCCCAGGAGGTGATATTATTTCAGT
GCTCAGCTGAAAATACCAACCCCAGGAATAAGAACTCCAAACAGTTCTGCATTCTGCGCCCATTCAAAATGTGTCTTTGTGCTTTGTGATTGCCTCTCCTCCCACTAGA
GGGGCTAAGCTTGACTGCCCTTAGCCAGCAGCAAGCAGGAGTGGCCCGAGCTCTCAGTGTAGTTATTTTTTTTTGTGTGTTGTTATGCCAGCATGGTGCAGCATGGTCTACATGGT
AGGAAATGAAAATTGCCTCTCAGTGACGTGACCAGGAGTGGCCCGAGCCTGATTTTTTTTTTTTTTTTTAACTGATAGAATGGTGCAGCATGTCTACATGGT
TGTTTGTTGCTAAACTTATATAAATGTGTGGTTTCAAATGTGTGGTTTTCAATTCAGCTTGAAAAATAATCTCACTACAGTACCATTATATGTAATGTTAGT
ATTTCTGCTTGAATCCTTGATATTGCAAATGAAATTCCTACTTTATTAATTGTATGTGAATTACTGTCTAGTTATTGTGCGATTTAAACTTTTTGCTTCTCCCT
TTTTTGGTTGTGTGCCTTTTTAACAACAGCCTCTTTTTTAACAAAGAAATAGATATTTAAAATTTAAATTGTGTGTGAGTAATGCTATGTTGTAATGCTAGTGTACTTAGGGAGTATGT
AAAATAATCATTTTAACAAAGAAATAGATATTTAAAATTTAAATTATTTCTACATAATGTGTAAAATGGGTAAATTTGGATTCAATGT
TTGTCTTTGGTTTTACAAGTAGCTTGTATTTTCAGTATTTCTACATAATGTGTAAAATGGGTAAATTTTCT
GACTTAAA

>HIPK4_H_SEQID#NA_25
GCACCCAGCCAAAAATAAATACATTTTCTAATTTGGCCTCAAACATCTCCCCAAGCTGCTTCTTATCCATGTCCATTTCTGGGGGACTGCCCCTTCATCCCAG
AGCCCTAGGCCGGTTCTTAGACAGTCCGGCCTTTCCTCAAAACCGTCGGCCTTTGAGTTGTCTCATCCATTGGCCTCGCCCT
CTCCTCCCCAACCTCTCCCATCTGTGCCCATGTCCTGGACTCTTTCCCTGGCTTGGGGCTCTTCCCACTCTTACCGGCCTCCAACCTATCCCTTTCATACAGTAACTTTT

>BIKE_H_SEQID#NA_26

TTTATGGAGGGAAACCCATCACCACCACCAAGGCTGATATCTGGCACTGGGATGTCTACTCTATAAACTTGTTCTTCACTCTTCCTTTGGTGAGAGTCAGGT
TGCTATCTGTGATGGCAACTTCACCATCCCAGACAATTCTCGTTACTCCCGTAACATACAATTGCTTAATAAGGTTCATGCTTAATAAGGTTCATGCTTAAT
CCTGATATATTTCAAGTGTCATATTTGCATTTAAATTTGCCAAAAAAGGATTTGTCCAAATAAAAAGCCAGAATAACAGTCCAGAATAACAGATAACAGAGA
GATGACTGCTAGTGAAGCAGCTGCTAGGAAAAGCCAACTCTGCTACTGCCACTCCAGTGTGCTGACCATTCAAAGTTCAAACACCTGTTAAAGTCCTGCCAC
CCAAAGGCCAACTCTGCTACTGCCACTCCAGTGTGCTGACCATTCAAAGTTCAAACACCTGTCCTGCCGTGGTGAATTCGGTAACCATA
GACCAAAAGGGGCACTAAGACCTGAAATGCCCTGAAATTTTATTGGGTCAGGGACCTCCTCAGCAGCAGCAGTAGTACTCCAGCAACTA
CAGCAGGGAGATTGGAGATTACAGCAACTCCATTTACAGCGTCATCCTCACCAGCAGCAGCCACCACTTCTCAGATGCTTATATGCAACAGCAGCAGC
AACAGCAACAGCAGCAGATGCTTCAACAACAATTTTAATGCATTCGGTATATCAACAACAGCAGCAGCCTTCTCACAGTATCCTACACAATGATGCGCAGTATCAGCAG
GCTTCTTTCAACAGCAGATGCTAGCTCAACAGCCATAATGCTCAACAACAGGCATCAACTTCTCAACCTTCTCACCAGTTCTCACCAGCCTTAGTTT
CCTACACTTCATCACTCCAGTTGGAACCAATAATGTCAGGGTGGAATCCTTTGGAGAGAATAGTTAACAGAAGAAGGGAACTATTGGACAGAGAA
ATCAGAAGAACATCAGCAATCCACCTGATATGTCGAGGAGCATCCTCGAAAAGAAAGCTGAACATTCATCTGAAAGTGATTTTGAAT
TTTGAACCTTCTAAGATCAAATATGGCTCGAGGAGCATCCTCGAAAAGAAAGCTGAAAAGGGAACAAGGAAGTTCTTCAGGGGAACAAGGAAGTTCTTCAGGGGAAC
GGTTCTGTTCCTTTCATTTCTCATTTCTCAAGATCTAAAGATCAGCGGACTGGAAGAGAGCAAGATGATGAAGAGAAGAGGCAAGAGATGAGGAAGAAGAGAGATGAGGAAGAAGAGAGATGA
AAAACAAGTCCAGCATCTAAAGATCTAAAGATCAGCGGACTGGAAGAGAGCAAGATGATGAAGAGCAAGAGATGAGGAAGAAGAGAGATGAGGAAGAAGAGAGA
CAGAATCCCCTTCTCCTAAGGCCTTCTCTACAGAGACAGAGCTCTCTTGAAGATCTGAAGTCAGTGCAGTGATGAGGAAGAAACCCAAGATCTTAATACAGGGGAACAAGGAAGTTCTTCAGGGGAACAAGGAAGTTCTTCAGGGGAACAAGGAAGTTCTTCAGGGGAAC
GAAAATCTGGTCATAGGCCTTCTCTACAGAGACAGAGCTCTCTTGAAGATCTGAAGTCAGTGCAGTGATGAGGAAGAAACCCAAGATCTTAATACAGGAAATCTTAATGATGAGCAGGCTAAAGCAAGTA
CAGTGACATAGGCCTTCTCTACAGAGACAGAGCTCTCTTGAAGATCTGAAGTCAGTGCAGTGATGAGGAAGAAACCCAAGATCTTAATACATAATACCTCTGATTATGAGCAGGCTAAAGCAAGTA
AGTGGAGACTCCAAACAGGAGTTTCCTGCTGCAGGACTGGAGCAGGAATTTGATGTATTGGGCGTGTCCCCTTCTTGCAGTGCGTGCTCAACAGGCGCCTTTAGCAGGCTCTCCACTCAATTATCCTCTGATGTTGC
TCAACACAGGTTCCTGCTGCAGGACTGGAGCAGGAATTTGATGTATTGGGCGTGTCCCCTTCTTGCAGTGCGTGCTCAACAGGCGCCTTTAGCAGGCTCTCCACTCAATTATCCTCTGATGTTGC
TGGGGCCTGAGGCACATACTATCCTGGTTATCCCAAAAAGTGAAATAACGGGGAGCCAAAAAGTAACGGGGAGCTCAAAAGTCAACAGTAAAAGTGAAA
GCAATGAGGACCTTTTGGGCTTGTGTGCCCTTTGATGAAATAACGGGGAGCCAAAAAGTAACGGGGAGCTTACAGAAACTGTCTCTGCCAA
AGGGCGCACAAAGCAGGATATGTCCAAAAAAGTGGGCGCCGAGACTCTCAAAGTAGCAATGAGACTTTGAAGCTACTATGCCACTCCAG
AGAGGGGCTCGCAGGCACAAAAAAGTGGGCGAATGTCTTACAACCTGGAGGAGAGAACAATTAGTGTTGCACTG
ACTGATGGGAAAGATAGGGGGAATGTTTGCTTACAACCTGCTGATCACAATTCTGGTGACCCCTTCGGTGACCCCTTCGGTGACCCCTTCATTCTCCAGACCCTGTTATGCGCACCCT
CCACATCAGGGCCTGAGCGACCATCCGTGCTGATCACAATTCTGGTGACCCCTTCGGTGACCCCTTCATTCTCCAGACCCTGTTATGCGCACCCT
TTGAAAATGGATGAATTTTGGTGCCCGTGCCCTTTACAGAACTTGTGTGCAAGCATCCACAACTCTGTTCAACCAGTCCAACCAGTCGCAGATGTA
TTTGGTTGCTGCTCCATTTCCTTCTAAACAGTAGATACTTCTGATGGAATTCTGGCATTCTCGGTTGATTTCTGATAGGTGATTTTGAAG
AAAATTTGGTAGCTCTTTATAGCATTCATTCTTAAAGATCAGTCAGAATAGGTGATTTTCATTCTTAAAGATCAGTCAGAATAGGTGATTTTCATTCTTAAAGATGAATGAATGAATGA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>NEK10_H_SEQID#NA_27
GAAGAATTGTCGGCATTGGCCTGGCCAGCGCTGGTTGGTGCGCCTGGACAGCGCTGGTCGGGCTGGTGGTGCGCCTGGCCAGCGCCGGTGTCCCTGAGAAA
AGTCAGAAATGGCCAATGAAGCTTTTGCTTATAAAGGAAATGCCGATGTTAATTCTGGGCATTGATTCTGGGCATTGAATGCCAGGTTGATCAAGGTGAAG
ACCACAGAAAAATCAACTGATAAAACAGCAAGAAATCACCATCAGCGACTATTCAATCTTTAAAAGACTTCGGTGCCTTTGAACGTGCAATCAAGCAAACA
ACAGCTTCCAGCATTAACTTGCCCAAAAATAGTCCAAATAGCATGAGCATGAACTACAAGAAAATTTCAGCAGCAAATTCAGCAGAGAAATTTCAGGAGATCTTTA
AATCCACAGAAGCTGTTGAAACTTGTTGAAACTCATAAGCAGCAGAGTGGGTTAATCGAGCCCATCATTTCTGAGGCTACTAATGAGGG
CCGCCTTGGTGAAAAATAGACTACTCCATAGCTTGGGTGGGGATTGAAAAACCTAGCTCAGTATATGGAGAATTGTAGCCAATGAGTTACCTCGGCTATGGAGAAGAG

```
CAGCACACTGTGGACAAGCTGGTCAACATGACACATATATTTTCAAAAACTTGCTGCAGTCAAAGATCAAAGAGAATGGTCACCACAAGTGGAGCCCACAA
GACATTAGTAAATTACTTGGTGCCGAGATACTAAGTGTTCTATTGGTTCCCTTCGGCTCTGCTAGTTAGCAGAAAGTCAAGAAGTTCAGAAGATGAGGAGAAGAT
AAGTGAACTCAACATTGTAGAAAATCTGTTGATGATTTTACATGAATATGACCGGTCCTCCTCAGTCTGCTCCACTCGACCACCTGAGCTCCTCTGAGCATTGTCTGGAT
GCCCCAGGTGAAAGAGCAGGTGAAGCTGAGACCCTGAGAAGCCTGGAAATTCGCAGGGCCATCAAACAGCTTCTTCATATTTACAAGGAGACAGAAATTTTGT
TCTGGTACAGGTTTGTGAGGAGCACTCCTCCATTGTGAAGCAGCCTGGAAATTCGCAGGGCCGAATCCAGCAGCTTCATTTATCAGAAGAGACTTGAGCCCTAGGGAAATACAAGA
AAATACTTTCTCACTTCAAGCAGCCTGTCGTGTCGCCCTCACTGAGCTGGTGCTCAATGACAAAAGTAATCTATTACAGTGTCTTATGTCTTCAGTGATGGA
AATAGCAAAATTAAATTTAACCAAATAAGCAAAAGAAATGCAGCAAAAGTACTTCATTGACAGTGTTTGAGAGCCTTGAGATTTCTCTCAGTGCTTATGAAGAATT
AAGAAAACAGACCACTCTTTAAAAGACTTTTCCCCACAGATCTTCAAAGACTGTGTTTGAGATGTGGAAGTGGGAGGATGAACTGAAAAGCTCCTTATTGACAGATATATAGGCA
GGTATCCAAGCTGAATTTTTGATCATCTTGGAAGGTGGAGGTGGGAAGGAGCTTTTGGCTGTGTTGTTTGTTTACAAGGTTAGAAGCCATATTAATCGAAAGCTCCTTGAAATATATAGGCA
ACTATGCAAGTCAATTTTGATCATCTTGGAAGGTGGAGGTTTTGGCTGTGTTGTTTACAAGGTTAGAAAGCATAGTGGTCAAAAATCTTTCTGAATTAACAATAATTAAAGAGCAGCTTTATCATCCC
TACATAACCCAGCATTTGGGAAGGGATAAGAAAGATCGAGACAGCAGCGTAAGGAGATCGAGACAGCAGCGTAAGGAGATATTGTTTCTGAATTAACAATAATTAAAGAGCAGCTTTATCATCCC
AACATTGTACGTTATTACAAAACATTTCTGGAAGAAAGACTATGAAATATTTATACAGCTGTTCTTCGTGATAGAAGAGGAGCCCGCTTGGAGAGCATTTCAGTTCTTTG
AAGGAAAAACATCACCATTTACTGAGAAAGACTATGGGGGATAAGGACAAAGTAACCGTTACACGTGTCTTCTTGATACTTTACACAGGAGGAGGAAGAAAACAGTAAACTCA
AGAGATCTGACACCAACAACATTTGGGGATAAGGACAAAGTAACCGTTAGCCGGTATGGGGAGAAGGCTGATGTCTGGGCAGTGTCTGGGCAGTGAGCCGTGATCTCTGGGCAGTGAGCCGTGATCCCTTTATC
CCTCTGTTGGTGGAACAATCCTGTATTCTTGCCCCGAGGTACTGAAGGAGTGGGGCTACAAAAATAGTGGAGGCGTATCGAGAGGAGCCGTGATATCGAGAGGAGCCAGAAGGTATCTTACT
CTGAAAAGTAACAGACACCATCAGCAGTGCCTCACTCTCACTGAGCTCACTCCTTGGCTACAAATAGTGGAGGCGTATATTGTGAAGCGTCAGATCGAGAGGAGCCCAGAAGGTATCTTACT
AATATTTAGACAAAACTTATCTACATCCCAGTTGTCCTTGGAAAAGAAGCTAGAAAGACGGAACGAAGAGCTAGGCAGAGGTATTTTATGGAAGCCAACCGGAA
CACCGTCACATGTCACCATGAGCTGGCTGTGTTCTATCTCACGAGACCTTTGAGGAGGAGCTGAGCAGGGCCTGTTGACGAAGACAGGGCCTGTTGACGAAATCTGTCAGATGATAACTTC
AACTTTCAGAAAAGCGCAGACCTGGCAGAAAAATGCTGAGAAAATGCTGAGAAAATTGGACATTTCGAGACATTCCGAGCTCAACGCCCTCTGAAAGAATC
AACCTGGAAAATGCTGAGAAAATGCTGAGAAAATACATATTTAAGTGTTTTAGTGTCTTCAGGAGAGGAAGAAAGACAAATCCAAACAAGGAACTTCACTGGAGGAAACAGGATCAAGACCA
TACATTCAACAGATGGCCACATTCTGTGTGCAAGCATCAGCAGCATCCTTCCTCCCAGCTTTCCTCCCAGCTTTGTGCCAGGAGGAAGTTGCGTGTCCCAGGAGGGAAATTGCGTCAGTGTCTACCAACATTTAATT
GGGCCACAGATGGCCACATTCTGTGTGCAAGCATCAGCAGCATCCTTCCTCCCAGCTTTCCTCCCAGCTTTGAAAGAAATCCCTCTTCAGCCAG
CAGCTGCACAAATAATCTATATCACACAGCTTCTATATCACACAGCTTGAAATCTGAAATTAAAAGTTATCTCAGGGATCTCCAGGATCTCCAGGAATTGAAAGAATCCCAGATTACCATTTATTAC
ATCGTTCATCCGGTGAAACAGCCTGTCCCAAACAGCCTGTCCCCAAATGACCCTGACCCTACCAACCAGCATTGAAATTGAGGAAGGAATAACATATGAACAGATGCAGACT
GTGATTGAAGAAGTCCTTGAGGAAAGTGGCATTTGAGTTGGACTTGATTTTCTCAGTGAAGTTCAAGTTCTGGACTTCTCGGAAGTTCTGGGACTTGGACTTGGGTGTGCTAGAGGG
AAAATGCTGCATTTGAGTTGGACTTGATTTTCTCAGTGAAGTTCAAGTTCTGGGGGTCCTGGGGGTCCTTGGGGTCCTTGGGGTCCTGGGGCTTCATTGCCAGA
TGTGGAAAAGACCAAGATGCCATGGGGCCTGCAGAGCTACCAGTGTGCTGCAAAAATGCTGCAAAAATGCTGCAAAATTCATCTTTTGGAGGGTGGGGGAAACCCAAAAACAACAA
CAAAAAAACTCTCTTACAGAATTTTCCTTAACATTAAAAAAACTTGTCATATTT
```

>pNEK5_H_SEQID#NA_28
```
CTAGCATAAAGAGGAAATAGAAGGAAAGCAATAAGTAAGAAAGTACATATTTCAATCTGAAAATGTTAGAGAAGT
AGCCAGTAGCCGCGCCCTGGGGAGTCGGGAGTCGCCGCCTGAACGTGACGGCGCGCTGAACGTGAGCGCGTG
GTTTTCCCAAATGGTGAACAATTCTGTTATCTGTGCCACAAGAAAGTTATTTGTGTCGTGTCTTGGCAAGGCTGGGAAGTTTAGCTAAGAAACTC
AGCCCATTGGACCATGAACTAAGTACGATGTGATTAAGGCCAAGTGCCTTCGGGAAAGCATACTTAGCTGAATCAGATAGCAGC
ACTGTGTCATAAAAGAGAATCAATTTGAACAAGGAAGAAGCTTCAAAGAATAAGAAGAAGTGATTCTTCTGGAAAAGAAATGAAACATCCCAAC
```

>NEK1_H_SEQID#NA_29

```
CTAAATATGGAATACCTTTAGCATATAAGAAAATATGGAGATAAAAAATTACACGAAAAATTAAACATAAACAGGCCCATCAAACTCCAGA
GAAGAGAGTGAATACTGGAGAAGAAATATCTGAGGAAGAAATATCTGAGGAAGCAGCAAGAAGGCTGGAATTTATTGAAAAAGAAACAAAAGG
ATCAGATTATTAGTTTAATGAAGTGTTAATGAAGGCTGAACAAATGAAAAGCAAGAAAAGGTTGGAAGAATAAATAGGGCCAGTGAACAGTATGAAACATG
TGCTAAGTGCTGGTGAAGTGGTGAAGTAAAGGCTCCTTTCTGGGCAGTGAGGGACTATAGCTCCATCATCTTTCTGTCGAGGACAGTGAACATT
ACCATGCCATTTTGACCAAATGCAGACAAGAGCAGAGTAAATGAAGCTAAATGAAAAGAGAAATATATGTCGAGTTCTTCCAGAAGAGGAAT
TCTGCCTGGAGTTCGTCCAGGATTTCCTTATGGGCTGCAGTGCATCACCATTTCCTGAGAAAGAGCTAAACAGTAGAAGAGTTCCTGCAGCGAAAACGGGAAGCTATGCAG
GTCTAAACAGCCAATGCAAACAGGCAAAAGGGCAGCTAGCTGTAGAAAACCTGCAAAACCTGCAAATAAGACTACACAGAATTCAATAAGACTAGATGGAGGCGAAATC
AATAAAGCTCGAGCGAAGGCCAATGCAAACAGGCAAAAGGGCAGCTAGCTGTAAAACCTGCAAATAAGACTACACAGAATTCAATAAGAGGCCAACAGATTAAAGCCAACTTCGTGGTGAAAAGAAAGA
AAGAGGAAGAGGTTTATCTGCAAGCAAGAAGAGAGAGGCTTATGGCAAGAAGAAAGGAGGCTTATGAGAGAGAATGTGGGAAGAGCAATTCGTGGGAAGAGCAATTCTGATAA
AGCTAATCATTCTGAAGGACAAGGACATATGGAATCCTGCAAAACCTGAGGCAATTCAATTCAATAAGACTACACAGAATTCAATAAGACTGAGGCCAACAGATTAAAGCC
CTAAAAGAACAACTAGAACGAAAGAAAGGAGAGGCTTATGAGAGAGAAAGGCAACATGAAAAAAGTGTGGAAGATCTGTTATTCTGTAAATAGCTTTGAAAGAAG
GTTTCTCCACCTTTGAACTGATACCCGGGAAAACTTCAGAAGAAGATGCTTCAGAAGATGCTCATCAAAGCAAAAGCAACATGCTATTTCAGTCTTTGTGTGGCGTG
GACAGTAGTTTAAACTGATAGATGAAAAAGGCAAGAAGTGGGAGGCAGGAGTCAACTTGAATTCCTGTCTAAAAACTGAAAACTGAAAACATACA
TCTTAAAGCTCAAGAAAGATGAAAAAAGCAGAAGATCTCTGATACTTTGAGATAAATGTTCATGAACATCCTTCTCTACAACTGAAAGACAATACA
GTTTCATCTGATCGCAAGAAGTGGCAAGAAGTGGGAGGCAGGAGTCAACTTGGATCTCCAAGAAGGCCTGGGCGAAAGTCCAGAGAAGCTGAACT
GTGGGAGAAGTGATTATTAGTTAGGTCCTAATGGATCTCCAAGAGGGGAAAAGTTCTCCGAGGGGAAAAGTGAGAATTCTCCGAGGGGAAAAAGTAC
ACAACTTCAGACGAACTATTAGAAAATACAACTATTAGAAGTGAAGTTACAAACCCTTAATTACTGGAGAAAAAAAGTAC
AATGTATTTCACATGAAATAAACCCATCAGAAATAAACCCATCATCCGTTGTGATTCTCCTGTTGAGACAAAAGTCCAGTTCAGTGAGCATCTCCACAGATGTCATTGAAAC
TGGAAGGAAATTTAGAGGAACCTGATGATTTGGAAACAGAAATTCTACAAGATGAACAACAAAGATGAGCTTGCCATGCCATGCATTACTGAT
GTGTGGATTAGTAGGGACAATCATATAGAGCCTAGGAAACAGAAACTCAGTGAACTTCAGGAGATCTCAGCCAATGATTCTCAAGTTCAGGAGATCTCGAGTACTG
TGGACCAACTAGTGACATTCAATAGAGCCTGGAACCAATGATTCTCAAGTTCAATCAGTTCAACTGGTCTGATTGGAACCATTTTCCATA
AGGTGGTTCATTCTGAACACTTGAACTTAGTCCCTAAGTTAGTCCCTCAAGTTAGTCCCTGATTGGAACTGGTGTTCAACTGGAACATCGCATTTACCACC
AAAAAATAAAACAAGAATTCCTTGCTGATTGGAACTGGAAGAAACTGACACAGATTTACAACTGCAAACAACCCAAAGATGTTAAGGACATGTTCACTTCCAGATCTCTC
AAAGCTGTTCAGAAGAACCCTTATGGATGTTCCCACCGTAGGGACACAGATTGCAAGAACAATCTTGAAGATGAAAAAACATTAAAGAAGGAC
CTTCTGATTCTGAAGAAGAAGAGCAGTCAGTCTTGAAGAAAACTGACACAGATTGAATGGGACAGAATGTGGCAGAATGAGAACAACCTGGTGAAGAA
TACAGTGAAGAAGAACGAAATGGCACTCAGATAACAGTGATGGTGAAATTGCTAGTGGTCTAGTGTAATGCGATAGTGTCTTTAACCATTAGAGGAACTGAGACTT
GTGCCCTGAACTGAAGAACGAAATGGGCTTTGAAAAATTCTTTGAGGTTTATGGAGAAAATAAAGGCTATTCATGAAGATGAAGATGAAATATTGAAATTTGTTCAA
CATCTCGGAGCAGGAAATGGGCTTTGAAAAATTCTTTGAGGTTTATGGAGAAAATAAAGGCTATTCATGAAGATGAAGATGAAATATTGAAATTTGTTCAA
AATAGTTCAAAATATTTGGGAAATGAACATCAGAATTCTTTATGCCAAGATTCTCATTAGTCAGAATAACAAGAGAGATAATGATGAATA
ATCCTCCAAAAATGTTTTTAATCCTCAACTATATGAAAGCATTTGGCTTATCAGAATAACAAGCCATTTAATATTATTTATTTAAAA
AATCAGATTTAAGATGGACTTTCTTATTGCATGAAAAAGATGGAGAAACATGCCATTTCAATGAAGATTCTAAATTTATCTATTTGTTCATTGAATTC
CATGGTTAAATCTCATAAAATATATACTTTATTATATACTTTATTAAATCATCCAAGAACCTGACTTAGGTGTCTGTTTCAAGTGTTTCAATGGTTTTGAAGATTTACTAT
GCAATAGGGTAACTTTGAGTTCAGCAAGTGTCTTTCAGCCATATGTCATTAAATCACCTATGCTGTCGTGTTTCAATGTTCAAGTGAGTCTTTAAGCATG
ATCTTTTTCTGTCTAGTACTTTGTTTCATTCTGGCCAGCAGTTCTACAATCACCGTGTTGTTACATCGACATTTTGAAGATGAAATT
TTTAGCCTTAAAGTTTATATTCTCAAGTCCTTTTTACAATGTCCTTTTTACTTAGTTGGTTTTTAAACTTCTTAGACTAGAATCATTGAAAAGATTTTG
ATTATGAAAGAATAGCAAAATTATATTCTGACTTAGCATATAAAAAAGTTGGTTTAATGCTCTTTTAACTTCTCCATATAAAGTTAAA
AAGGAGGAAAATAATATTTATTATTTTTATTTGACACTGTGACAGTTTTTGGTAACTAGATCTGTTGAATGTTTGCCTTGTTGAACTTCTTTCTGTTATGAGAG
GATTTAGTTAGGTCATTAAGATGTGATCAATCAATCACAAGCTTCAATCACAATATAACCTGTTTCGTTAGAGGTGTCTCACAGTCCAGTGTCTTCGT
```

AATAAAAGCAAAGTTTTGAACCTCTGAGTCAGTTGGCATAATATGTAATTTGAAAAATAAAATCTTATCTTGCAGCACTATCAGTATGTT
GAATTTATTATGTATATTATTTCTAATATCCGAAAATCTGATTTTTTAATATGTGTTATTTATTTATGATATTGCTATTAAATTTTATTATCT

>NEK3_H_SEQID#NA_30
CAGAACTCAGTTAGTCGGGGACAATTTCCCTCAATGTTAACAGCACTGTTCCACCGCAAGTGGAACAACAGCTTTAAAACGTGCTCTTCGTAGGCCCGGCT
ACTCCAAGAACAGTGCCTCCCGCCAGACCCAGCGGCGGCTTCCTCACCCGAGAGACGACCCGCCAACCCGGGAGCCCGCCGGTTGCCA
GGCCAAGGAGTGGACTAGGCTGCCGCGGGAAGCGGTTTGGGAGAGCCGTCCTTGGGCAGAGCTCTTTGGTTCAGCAGCGTGAGTGTGTTGCCTCATGAAGATGA
CTACATGGTCCTGAGAATGATTGGGAGGAGGCTGCTTCTAGGAAGGAGGCTGTTTCTAATACACAGAATTCTAGAGATTGGAATGAATCATTTGAAGCT
TCCCAAGTCTTTCTAATACACAGAATTCTAGAGATTCTAGAGGCTGTTTCTTTAGCCAAATGAAAACACCCTATATTGTTGCCTTCAAAGAATCATTTGAAGCT
GAAGGACACTTGTATATTGTGATGGAATACTGTGGTAAATCACATTCACAAGAAACGTGTACAAGAAAGGAAGTTATTCCTGAGACATGATACT
TAATTGGTTTACCCAAATGTGCCTTGGAGTAAATCGATGGCATTTGCTTACCATGTCCAAGAATATCTTCCTCACTCAGAATGG
AAAAGTGAAATTGGGAGACTTTGGATCTGCCCGTCTTCTCTCCAATCGGTTCATGTCATCTCTGTACCCTTAAGCATCCATTCAGGCAAATAGTTGG
AAAATCTTATCCTCAAAGTATGTCAAGGGTGCATCGAGTCCACTCGTGAGCCTTACCCCCGAGATCATCATGGAATCATGGTGAGG
CCTCACATGGCGCCCCTGCTACAACGCTTTCTCTCTCGAGGCATAAAAAAACACACCAAGAAATCAGCAGAACTCAGTCAAGCAGATGTTAAAGGAATC
AAGTATTAGAAGAAAATAAATTCGAAGCATAAAAGGGTTAGCCCATACCGATTTGGAAAGTCATTTAGTTGAAAGAGAAGAA
GCAAGAGGAAGAACAAGATAGAAAGCCAGTTCACATTCCAGAAAAGCCAGTGGGGAGAAAAATGTACTACTCGTTTGGA
AAAAGTAATAAGTCATCCATAACTCCAGTTTAACAGCAGACGATAAGTGCTGATTCTCAGCTTGAACATCCTTAAGGTTACAGTGGCTCAAAG
AGACCCCTGACACTTTGTTGAACAACAGAAATGCTAAGATCCTTAAGCATCGGACAGTGTTGATGGAGGTCACGATCTGAGCCTGGGCTAGATGAGGAG
GCCCCCGTGTCTGAAGAAACAGAAAGAAGATGACAACCCGACTGGGGTGTCAGGATGCTCAAGGCCTGTGCGACAGATAATGCCTGAGGA
GACAGGACTTTGAGGAGGAAGAGATGACAACCGACCTTCACTCAGGAGTTCATGCTGAGATGATCATGCGGACGTATATTTCCTTTGAAAACAGAATGAAG
AATGTTCCTGAGTCACGCTGAGGAGAGCTCTAATAATCTTAAATCGTTCTTGATTAGTATCGTGAGTTTGAAAAGTCTAGAAACTCAGCATTAACCTCT
GGAATGAGTGTGAGCATCGGGCTTTGCAGTCCATAAAATGAACAGAAATGCCTACCTACTTGTTTGATTGTGGGAAATTACTTAACCTCT
TCAAGCCCCAATTTCCTCAACCATAAATGAAGATAATAATGCCTACCTCAGAGGGATGCTACCTCAGAGGGATGCTCGACCACCAGACCTTATAGCAGCCGTATGATATTATTCCACA
TTATGATATGTGTTTATTTATTATGTGACTCTTTTATTATTTGTGACTCTTTTATTATTATGTGAATATATATTTCCTAAAGGTTTGAGAATTAAATATATTTAATTATGAAAAAAAAAAAAAAAA

>SGK069_H_SEQID#NA_31
AGTCGGGGCGGGGTCTTGCTCCTCAGGCAGGCCCTCTGCTGGCATGAGCCCTAAGTGCCGGCGGCCACTGACCAGCCGCGGGAGGGTCAGGAGGGCCTTG
GAGGAGAGATGCCCGGCAAACAGTCTGAGGAAGGGCCGGCGGCAGGGCGGCCTTCGGAGGAGAGGGTCTGGGCGGCCTGACATTAGAGG
AGCTCCAGCAGGGCCAGGAGGCTGCCCGCGCTGAGGAGGACAATGAGCGCTCGAGTGCTCAGAAAGGCGGTGACAGCTCTACGAGGA
AGTGCGTCCCTCGGCGTCCTTAGCCCATGGCCCTATGCCGCGGTCGTTCGTGTGGGCTCTCGCACGGACTCGAAACCCCGCAC
GTCCCTCCGTGGCTTCCTGTACGGAGTTCGTGTACGGGACCCGTCCTGCAGTTGGCATCAGCCTAGCGGTTGCATCGGGCAGGCGGTGCACACTC
CTACAGCTTCCTGACGGAGCCCGTCCTGCAGTTGGCATCCAAGTGGGGCTCCCGAGGTGGCCGGTGCACCGGCGGTGCCGCGCCA
GCTGGCCTCCGCCCGACCCTGGTGTGCACGCCGCTGGCGGCCTGGCGGCCTGCGGGCGGCCTTCGGCGCGCCCCGGCGCCCCTGGC
AGCTGACCGACTTCGCGCCACCGCCCGCCATTCAGCCGCGGCCTCGGCGGTCCTGGCGTCCTGGCGTCTTCCCTGGCTACTTCCCTCTCACGCGGGACTCCTGCAGCCCCTGCAGCCCCTGCAGCCCCTG
CGAGGGCCGACCCCTTCACGAGGAGGACTTCCTACGCAGGAGGACTCCTCCATCTCGCAGCCCCTCAGCCCCTGCTCGGGCCGCCGCGGCCGCACGC

GCTTCTGCGGGGGCTGCTGGACCCTCACCCCCGAAGGAGGAGAGCGCGCTGTGATGCGCCATCAGGAGGACCTGGGCGCGGGCCCCTGGAGCGAGCGGGGAGGGGCGAG
GCGGGAGGCAGTGGGGAGGCGGTGGAAGAGAGGAGGGCTGGGCAGTGA

>SGK110_H_SEQID#NA_32
GCTTTCTGGGGAGCCCTGACGGGGGGGGTCCCGGCCCCAAGGGCCACCCCACAGGCGCCAGGGCCTCCGAGACCCCTGAGGATGG
GAACCAGAGGTGAGAGAGACGGGGAGGAGGCAGGGGAGGCAAGAAGACCAGAGGCAACACACAGCGACGAGAGGAGGAAGAGA
AACAGAAAACAGGGAAAACGCAAAGACCTCAGAGACGGGGATGACAGAGATCCAAGGACACACACAGCGAGACCAAGCCCCAGAG
AAATGGACTCAGAAAACACACAGCGGGATGCACAGGGAAGAGAGACCAGGTGGCAGGGGACTGTGTGGCTGAGGATGCACC
CCCACTCGGGGCTCGGGGAGGCTCCTGGCCTGTCCCACAGACGCCACAGCCACCAGCCCTCAACCTCTCCAGGAGCAAAAACAGGTG
GCCCGAGTCCGGGAGGACACAGCCACGAGCCACCAAACGGCTGTGTGAGCTGACGACCAGCAGGTGACCCGGTTCGGGACCAGTACCACC
TCATCCGGAAGCTGGCTCCGCTCCTGAGGGGAGTTCTGTGTGGGCGGTCCAGCCTCACCACCCAGGGGGTCTCGTCGGGATTTGG
TCCTGAGAAGCACCTTCGCCCAGAGTACGCGCCCTGTGTGGGGACCTCAGCGGGGAAAAGGGGCCTCCAGAAACTGTCTGGTCTTCGCAGCCGTGTGGCC
ATTTTGCCTTCGCCCAGGAGCTCTGGACTTCCTCCACAGCGGCCTGGTCTGACCCGGCAGCCCGAGCGGCAGCCCGTGGGTGCTGCTCCTACCGCCC
CTGGGAGACCTGGGTCTGACCCGGCCAGCGCCTGGCTCTGCGCCGTGGAGTCCTGGGGGCTTCGCCAGCGGCAGCAGGGA
GACACCCTGAGTTCGGAGGCCTTGCGCTGGTGGTGACCTGAGATCCGAGAAGCTCCTAGGGACGTGGCACTGGCCACCTG
ACCCTGAGTTCGAGGCTTCTGCGCTGGTGGTGACCTGAGATCCGAGACTAGGAGCACAGGTGTCCTGATCCGAGACTAGGAGG
TCCAGGGGCTTTGGGGACAGATGGGGGAGCTCCCTGACCAGGTGTCTGATCCGAGACTAGGAGCGGGGACAGATGGGGGAG
GGAGGACGGGGACAGAGTGCACCCGGAGGAGGAGAGACGCGCCAGGATGCGGAGGAGCTCCTCTGTTGCCAATCTTCCAACTGAGGAC
CCAAGAGCCCAGCTCTGCCTGCCCTCTCTCCATCCTTCCCTCCCGTCCAGAGTGTCTTC
TTTTTAITTTTTAITTTTTTAGACAGAATCTCGGTC

>NRBP2_H_SEQID#NA_33
GGGGGCGGGGGCTGGGCGGGGCGGGGGGCTCGGCAGCTCTGCTCTGGCTCTCCGGGCAGGCACAGGCGGCTCGGGAGGTGCGCGCCGGCTCCG
CCCCCGGGGGGGCCGGTCCGCGCGGCCGGTGGGGGCTCCCCAATTCCCAGAGCCGGCGGCTCTAGGCTGCGCGGCGGTCGGGGCTGCTGTGGG
CGCGAGGCCAGGGGCCCGGGAGAGGCTGCGGGATGGGCGGCGCCGGCGGAGACGCGCAGAGGCGGGGTGCCCGGCTGGCCGGAACCATGCGGGCCCCGGAGCCGGCGC
CGAGGCGGGGGCCGGGAACGGGGACGGGAACGGGAGGAGGCGGGAGGACGAGAGCATCTGAGGAAAGCCCGTTGGTCGTGCTGGCAAAAGCGAC
GGGAGCAGGTAAACCAAGGAAGGCCTTCGCGGCGTAAACCAAGGGAGAGAATGGCAGGGCTTCAGAGAGAACATGCCAGGGCTTCAGAGAGAAGCAACGGACTCCACTT
CGGAGAACAGGAGAAGGCCTTCGCGGCGCACGAGGAGAGAGATCCAGACCGTGTTCAGAGTACGTGTTCATCAGAGTACGTGTTGGTGTGGTGACCGGAACATGTGAAGTTGCACAAG
TACTGGCTGGATACCTCTGAGGCCTGCGCGGAGGGTCATCTTCATCACAGAGTACGTGTTCATCAGAGTCGCGCGAGGGTCATCTTCAAAAAGACCAAGAAGAA
CCACAAGGCCATGAACGACCACCATCTTCATTCAGCACAACGGCCTCGTGTTCGCGCTCAGGCTTCCTCCGCGCCCCCAATCATCACGGGA
ACCTGACCCAGCGACCACCATCTTCATTCAGCACAACGGCCTCATTCGGAACTTCTCCAATGTCTCGCCCTCCCACAGCAC
TTCCAGATGATCTCCAGTAGCTCCTTTGGGATGTGTGCGCTGGACCCCAACATGGGACCCGCGTCCCAGTATGGGACCGGCGCTGCCATGGACCGCT
GTGGACATCTTCCTTTGCACTCGCTGAGTTGTGCGCTCTGACCCAACATGGGACCGGCGTCCCAGTATGGGACCGGCATTGCTGCGC
CAGGCAGGTAAACCAAGGAGTAACCAAGGGAGTTCATCTTCGCTGCGAAGCTCTGACAAGCTCTGAAGCTCTGCCTTCCACGGC
GTGCTCTTCGAGGTGCACTCGTCAAGCTCTGCCTCATCATGCGCGCTCTGCCTCATCATGCGCGCTCTGACCATGTGGTGGAGAAGACCAAGGCC
ATGGACCTGCACGCGCGTCTTGGCGGAGCTTCCCGGCGGAGCTTCCGGGGGGTACTCGCAGGTGGCGGTACTCGCAGAGTCGTCCTTCATGAGCGGTACTCGGGTACTCGCAGAGTCGTCCTTCATGAGCGGTACTCGGGTACTCGCAGAATTCCTG

GAGGATGTCAGGAATGAATCTACCCACTGATGAACTTTGCAGCCACTCGACCCCGTGTGCTGGCCCCGTGGGGCTGCCCCGGAGGAGGTCCAAAA
GGCCAAGAGACCCGAGCGCCAGAGCCCTTTGACTCTGAAGCCCTTTGACTCTGAGACCAGAAAGGTCATCCAGATGCAGTCGAGTGCAACTGGAGGAAGGCGAGGAAGGCGCGCTGGCAT
CTCACTCTGCTTCTGGTGCTTCCTCCACGAGGACGACCGGCTGCACCGGACCTGACCTACGACCGGCTGCTCCCAAACGACAGCGCCCAGGCCTGACCCGGAGCTCGTGTGCAC
TATGGCTTCCTCCACGAGGACGACCGGGATGAAGCTGCGCCGGATGAAGCTGCGCCGCCTTCCTGGAGAGCACTCCAGTACCGTGGCCATGTGTGGGCAGGCCATGTGTGGGCAGGCCACAAAGCCC
CCCAGGGGACCATGCGCGGGGTGCTGCCTGGGCAGGCCATGTGTGGGCAGGCCGTTGGGCGACCCTGGGGCGCCGGAGGGTTGGGGAGCCGGTTAGGCCTCCCAG
CGGTAGTGAAGGAACCCCCGTCTCCTGAGAGTGGGCGTTGGGCGACCCTGGGGCGCCGGAGGGTTGGGGAGCCGGTTAGGCCTCCAGCTGCCTCCTCAGC
GTCCTTAGGATCAGGGTTGCCCCAGAACCCTTCCATATCTCCGCCCAGGTTCCGCCCAGGTTCCGAGTTCTACCAGGTCCATTCTCCGCCCAGGTTCCATGGGGTCTGG
ATGCAGGAGGCTGCCGTGTAGGGAACCCAGCTCTGGGGTCTTGGGGTCTGGGGAACTCGTTGCACTGTGCACTGTCAGGCACTGGGCA
GAGAGCAGCCATCCCTGCTGCAGGAGAAACAGACCACACAGAGAGGTCTGGAGTTGAGGTCTGGAGTTGAGGCTTGTCAGGCAGGTGTCAGGCACTGTGTCAGACTGAGCCACC
CAACAGGGACCTGGCAGGAGAAACAGACCACACAGAGAGGTCTGGAGTTGAGGTCTGGAGTTGAGGCTTGTCAGGCAGGTGTCAGGCACTGTGTCAGACTGAGCCACC
TCTTTGACCCCTTTTACCCACCCCTGAGAACCTGAGACCCCAACAGGGTCCAGGGTCCAGGATCTCCTGCAGATCTCCTGCAGATGGGCTTTTCTCTCATGCATTCCC
TGGCCTGGAGGGCGTCAGGGACTCCTGCCTCCTCCCACTCACGGGACCTCCCGCACCTCCTCCTGGCAGCTGCTCCCTGGG
CCTCTGAGGATGTCAGCTCTTGGCTTGCCTCTCTCCCACTCAGTGAAATGGCCTTAGCATTTTATACTGTGCCTTAGCATTTTATACTGTGCCTTAGCAATTCATGCCTCCTTCCTG
ACGCATTTCAAAGTCTGGAAGTCTGACAGTCTCTCACTGGGTCTCACTGGTCTCACTGGGTCTCTAGATTCTATGCCCTCATTCTATGCCCTCATTCTACCCGTCCTTCTAA
AATTCTGTAGGCTGAGAGTCTGCAGAATCCACATCCTTTCGGTGCCCCCTCCTCCATCTTCGGTGCCCCCTCCTCCATCTTCTCAGAGCTAGCTTAAGGGGCCAGCTTAAGGGGACTCTGTCTCCT
GCTCCTTCAGGCTGCTGCAGAATCCACATCCTTTCGGTGCCCCCTCCTCCATCTTCGGTGCCCCCTCCTCCATCTTCTCAGAGCTAGCTTAAGGGGCCAGCTTAAGGGGACTCTGTCTCCT
GTTGTCCCTAAGGACTCATGAGATTAGGTTGGGCCCCTAAGGACTCATGAGATTAGGTTGGGCCAGGATTAAGGAACGCATATAACATGATAATCTCCCTCCTCAAGGTTTTTAATATTAAACACATCTGAGGACACA
AGCCAAAAACCAAGCAAGACTTACTAAGCAAGACTTACTAACGACATATCAATAAATAGCAATCTCAAATTAGCAATCTCAAATTAGCAATGTGTGTCTGATTATGCATGGTT
TTATTCAAATTCTCACGCCTTCAACTAACGACATATCAATAAATAGCAATCTCAAATTAGCAATCTCAAATTAGCAATGTGTGTCTGATTATGCATGGTT
CTAAATGGTTTCAGTGAACTGAAACATGGTCAAATACATAACATTGTACTACTGTTAAACTGAACTTAAAGCAAAAAATCTATGTCGGGTGGGAGAGGTAATGAAA
TGGCA

CCCGGGCAAAAGCGAGCGGCGCCGCCCCTGCCTCTCCGCTGCTGGCAACGCTGATCTATCTAGTTGCTGGGGAGACGCCCCAGATGCCGGGCCCACTCG
GACTTCAGCACACATCCGAAGGATGGGAAGAAAGAGGCCCCACGAGCGGGACTCGCAGGCGGGCCCCAGAGGCGGGGTGAGAGGCGGGGATCAGCT
GGCCCCTTGCGCCTGGTTGCACTTGCATGTGACTAGCTGCGCGGCGCCCCGGGGTCTGGCGGTCTGGCGGTTGGTGGGGAG
GAGCTTCTGGGTGATGTAAGGCCGGGAATGGAGGTGGCCTCTCCTGCTGCTCGACTCGCTGCTAGGAGGGGCGGGACTCGCAGGAAGGGGCCGGGCGCCGGGGAGGGG
GCAGGCGGTTCATTGATGAATAAAACGCTGGGGCTGGGCGCCGGCCGCCTGCTGGCCGCCTGCTGGCGGGGTTTCCTGTCTCCCGCGCCCTTCCTGTCTCCCGCGCCGGGAACCGAGA
AGCCGGGGACCCCGGGGGCCTCCGAGTCGTGTTGGGCATGTGGGGGCTTCCGGGGCGGGGCCGCTCGGGGCTTCCGGGGCGGGGCCGCTCGGGGCCGGGGCGGGGCCGCTACGGACCCGGCGCA
CGGGGCCCGGGGGCTCCGAGTCGTGTTGGGCATGTGGGGGCTTCCGGGGCGGGGCCGCTCGGGGCTTCCGGGGCGGGGCCGCTCGGGGCCGGGGCGGGGCCGCTACGGACCCGGCGCA
GCATCCCGACAGAGAGCCGGTCGCCAAGCCGGTCATCAGCCGGCGCTCCAGCCGCATCTGCAAGCGCCTCAAGTTGGGAAATTTTTCAT
TTCGGCCCACCTTGGAGTGCGCTACTCACCTGCGCTACTCACCTGAAGGTGGGGGAGATTTTGGGGCTGGAGCCTGGGCCGGTTGGAGCCTGGGCCGGTTGGAGCACCCCAACTATG
TGGCTCCAGAAGTGCTGCTGAGACAGGGCCGATGTAGGCCGGATGTAGGCCGGGCGCTGTTCTGTCGGCGGAGCCCTTGTCTGTCGGCGGAGCCCTTG

>SCYL2_H SEQID#NA_35

ACAAACCAGAACTTCTACAGTAGTCCAAGCACAGTTGGAGTGACCAAGATGACTCTGGGAACACCTCCCACTTGCCAAACTTCAATGTCTTGAGTGTTCCT
CCTGCTGGTGCAAAGCAGACCAACAAAGACCACAAGACCCACAGATATGTCTGCCCTTAATAATCTCTTTGGCCCTTCAGAAAACCCAAGTTAGCATGAACCAGTTATCA
CAACAGAAACCAAATCAGTAGTGGCTTAATCAGTTCAGTTGTACCTCCTCAAGGTTCTCCAAACTTGCACAGCCACCAACTACTATGGGCAGTTCAGTCAGTCAATGATTTAAA
ATCTGCTTTTGGTATGCAGGGTGTCTTACTTCTATTTGAAGGATTATTTCAGTTCAATCATCATGGTGAGCTGATTTACATCTTTATATAGTTGGCTTGGAGGAAGTA
GATCTTTTGGGTGAGTGTCAGTTCTGTGACAGGAAACATCTCTGTCCATGCCAGCATAGTAGTGTGTATGATGAGTTGGAGTTTTAAAGCATTGA
CTTCTATGGGAAAGTGAAACAGTTCTGTGACAGGAAACTCCTCTTCAGGTTTTAAAGACCCAGCCCTTCCAATCTCAAAGAGAAAAGGAAACTGAGTTATCTTGAATAACATAACTT
GGATTTTTCCTCTTACCAACTCCTCTTCAGGTTTTTGGCTTGTGGATCTGGTTATTTAAAAATTGAGGTGATGGTCATTGCAAGCTCATCTTATTAAGTACTATATGGTACACAGTC
TTTAATCAAATGTTTATTTGGCTTGTGGATCTGGTTATTTAAAAATTGAGGTGATGGTCATTGCAAGCTCATCTTATTAAGTACTATATGGTACACAGTC
TATGAGTCATTAGTCTTCATTTTAATATGTAAAAAATCTTGATGCTATTGTTTAGTACAGAGCTGAGCAAATACTACAAATTCAACTTCATTTCATTCATTGGTT
GTATCAGGTTATTTGCTTTTTCCAAACTCTTAAGTGCAAATCATTGTAAATTATATTTAGCATGGTCTGCCTCAAATAGTAATGTTTCTGCATTCATTCACTTGGATA
TAAATGCGTTATTAAACCATCTTAAGTGCAAATCATTGTAAATTATATTTAGCATGGTCTGCCTCAAATAGTAATGTTTCTGCATTCATTCACTTGGATA
TATTTAGAAATCACTTTTTCCCTCGTATCAAGGAAGAGGTATGTGCTGATTTGTTGTTGGATATTTGTTCAAAGAAGATTTGTTGCATTTGGATATTGTGCATATAAACTATATTAACTATATTATGAGTAAACTTTT
CATATTTCATTTCAAATTCAAACTTCTGAGGTTGCAGCATATGAAATTGAGCAGTTGTATCTTCTGACCACCAACAGATTTTCAGCTTGCCATGATAG
GAGGTTCTGCTGTATTGTTCAAATGTAATAAACTTTACTTCTGTAATAAACTTTCAATTTTTTCTGAAGTTCAATTTTTTCTGTTAAATAATTTTAGGAAGGA
TCTGACCTCATTAATAACTGTCTACTGGAAACATGCTTTATTTTCTTAAAAAACCTTTCTCTGATATTTTCTTAATTTGCTGATTATTCAACCACAGAGCCTTATGCTATA
AAGAATCAATTCTCTTAACAGGAAACATGCTTTATTTTCTTAAAAAACTTTAAAACTTTAAATAAACCATCTTAAATAACTTTAGTTATTAGTTGTGCTGCCATAGAGTTGATGG
AATGTCATTTGTATTTGGAACCAGCAATCACTTTAAAAATAAACCATATTAGTTATTGCTTGGTATTGTTATTCATTTTATGAAATATTCATTGAAAATATATA
TTTTAATTATCTGGAACCAGCAATCACTTTAAAAATAAACCATATTAGTTATTGCTTGGTATTGTTATTCATTTTATGAAATATTCATTGAAAATATATA
CACAATATATGTAATACACAGCACTTGATTACAAAGGACTAACAGAATTGCTAAAAGGACTAACAGAATTGCTAAAAGGACTAACAGAATTGCTAAAAGACTAAAAAACAG
CTTTCACATCTGATTTCTGTATGGGCTGCTGTATGGGCTGTAACTGTTAACTGTTAACTGGTTAACTTGTTAACTGGTTAACTGGTTAAGATCCTACAT
TAGGAAGATCAAATGTTTTGTCAAATAGTTTTGTCAAATAATTTTCTTGTTAACAATTTGTAACAATTAGTAGTCACTGTTTATGAGAAATTGTTTTTTATTTGTAAAATAACATGATGTTAGTGT
ATGGAATAAAATAAGTATTGAAGAAATTGAAGAAATGTTTTGTAACAATTTGTAACAATTAGTAGTCACTGTTTATGAGAAATTGTTTTTTATTTGTAAAATAACAGTATGTAAATT
TGAACTCTTAAACAGAAAGAAAGCTTAATATAACAGCTTATAGAACTTGAACTGAACTGAAATATGAAAAATACAGTTATTCATCCATCAATTGAACTGTT
TGTTCATTCGTTGAGGTAATGTCTGTTGAGAAATTGCCTCTGAAAAATAGTGCTATTTGCCTCTGAAAAATAGTGCTATTTTACTTTTTCAGTGTCTTAAGTGAATGAAATTTCAAAGTGAATGAATCTCAATGCTACTCACCTTA
ACCATTGCCTTTTCTGTTGAGAAATTGCCTCTGAAAAATAGTGCTATTTGCCTTTAGAGTAGAAAATGTTAGGTTAAAGAGCATCGTCAACAGAATCTACAAAAAGATTCCC
AAATTATTTCAGTACTGAAGACAATTAAGTCCGTTATGTTGCTATTGCTAAATGTGTGTAAATGTGTGAATACTGAATATAGAGAAGGATGTATAAAAGGAAGTAGTATAGACTATGTACTTGTCTGGTTTTT
TTGCATTTGAATTAGTTCTCTATTTCTCTATTGCTTAATTGTGTGTAAATGTGTGAATACTGAATATAGAGAAGGATGTATAAAAGGAAGTAGTATAGACTATGTACTTGTCTGGTTTTT
GTTTGTTTTTTATTTTGGAATGCTTATATGCGAACATACCTGATATCAAGAGGAGTATGCACCAAATAAATTTAGCTTTGATAAAACTTTCCAT
CATGAAAGAAAAAGTATATGCGAACATACCTGATATCAAGAGGAGTATGCACCAAATAAATTTAGCTTTGATAAAACTTTCCAT

>SRPK2_H_SEQID#NA_36
ATAGCACCATTGAATCCCAGTCCTAACAGAAGTACTGCGAATCTTGTGGCCTCATTCTGAACAAAAGGGATTAGAGAAGAAAATCTCTGATATAAGGCTT
GAAAGCAAGGGCAGGCAATCTTGGTTGTGAATATTTCTGATTTTTCCAGAAATCAAGCAGAAGATTGAGCTGCTGATGTCAGTTAACTCTGAGAAGTCGTC
CTCTTCAGAAAAGGCCGGAGCCTCAACAGAAAGCTCCTTAGTTCCTCCTCCCACCACCGCCACCTTTGCCAGACCCCACCACCCCC
GGAGCCAGAGGAGGAGAGATCCTGGGATCAGATCAGATGAGGAGCAAGAGGACCCTGCAAAGGTGGATATCATCCAGTGAAAATTGGAGACCTC
TTCAATGGCCGGTATCATGTTATTAGAAGCTTGGATGGGGGCACTTCTCTACTGTCTGCGCTGTCGGGATATGCAGGGGAAAAAGATTTGTTGTGCAATGAAA
GTTGTAAAAAGTGCCCAGCATTATACGGAGACACGCCTTGGACATTTCAGCATGAATGGGGATACATGTGTCAGTCTCTTGCGCACCATCTCCTCAAGTGGATCAT
GGTGGTCCAGCTCATTGACGACTTCAGGATTTCAGCATGAATGGGGATACATGTGTCAGTCTCTTGCGCACCATCTCCTCAAGTGGATCAT
CAAATCCAACTATCCAAGGCCTCCCAGTACGTTGTGTGAAGAGTATCATTCGACAGGTTCCTTCAAGGGTTAGATTACTTACACAGTAAGTGCAAGATCAATTCATTCA

TACTGACATAAAGCCGGAAAAATATCTTGATGTGTGTGGATGATGCATATGTGAGAAGAAATGGCAGCTGAGGCCACTGAGGTGGCAGAAAAGCAGGTGCTCCTC
CTCCTTCAGGGTGTCTGCAGTGAGTACGGCTCCACGACAGAGAAACCTATAGGAAAAATATCTAAAAACAAAGAAAACTGAAAAAGAAAACAGAAGAGGCA
GGCTGAGTTATTGGAGAAGCGCCTGCAGGAGATAGAAGAATTGGAGCGAGAAGCTGAAACATCACCTCACGTCGAACAATGGTGAAGCT
AATGACCAGGATGGCGAATACTGCCCAGAGGTGAAACTAAAAACAACAGGATTGAGAGAGACTGCAAAGGACAATAGACCCTACGTGGATA
GAGGACCAGGAAGAAGAAGATGCTGAGAAAGATGAAGATGATCAGGACGATGAGAAGATGATGAGACGACATAAAATTC
GAATCACCTAAAACCAATGGCCATATTGAGAATGGCCAATGCAAATGCAGAAAGTGATTACACATATGACACATATAGCAGCTCCTATGAACAATTCAATGGTCAAATTGCCAAATGGACGACATAAAATTC
GGAATATAATCTTGATGAGCCAGAGTTTCCCCAGAGTTTCTGATCCTTAGAACCTTGGCGCCTCTGTGGCCTGGCTTTCTGAGGGATCAACCACTTACTGAGCA
AGAGGAGCAGCAGTCCATGCGGAATGCAGATATAAAATTAGAGTAAAAATTGCTGACCTGGGAAATGCTTGTGTGGGTGCATAAACACTTCACGGAAGACATCCAGACG
ATCCCCTGGATCCGCGGAATGCAGATAGAGGTTTAATAGGGACGGGGGTACAGCACTATTCCAGAGACGAGAACTGCGACACACATCACCAAGCTGCTAGGCACTTTGCTCT
CGTCAGTGACGGCTCCATAGAGAGTTTAATAGGGACGGGGGTACAGCACTATTCCAGAGACGAGAACTGCGACACACATCACCAAGCTGCTAGGCACTTTGCTCT
TTATTTGTTGAACCACATTCTCGGGAATTCTTCAATGCAGAAGATTCTTCAGCAGCAGTTCCTGATCCCGATGTTAGAAATGTTCAGCAGTATTCTTGATGTACTTGTGGAAAAGTA
ATCTGAAAAATATTCTCGGGAATTCTTCAATGCAGAAGATTCTTCAGCAGCAGTTCCTGATCCCGATGTTAGAAATGTTCAGCAGTATTCTTGATGTACTTGTGGAAAAGTA
TGGCTGGCCCCCATGAAGATGCTGCACAGATTTCCTGATCCGATGTTACAGATTTCCTGATCCCGATGTTAGAAATGTTCCAGAAAAACGAGCCTCAGCTGGCGAATGCCTTCGGCA
TCCTTGGTTGAATTCTTAGCAAATTCTACCAATATTCTGAGCTAGCAAAGTGTCCAGTACATTGGACTAGCAAAGGTGACTCTCATTCTTTAACAGGA
TTACAAGTGAGCTGGCTTCATCCTCAGACCTTTATTTGCTTTGAGGTACTGTTGTTTGCTTTTGCTCTGTGTGTCGTTGTGTGACATTCAATTTTTTTGA
TTGTCTTCAGCTAAGTAGTTACTGACCATTTCTTCTGGAAACAATAACATGTCTTCAAGCAATGTTTCTGTTGTGTGTGACATTCAATTTTTTTGA
ATGAAAAAATACTTTCCCCTTTGTGTTTGGCAGTTTGGCAGTTTGAGTAGTCCTGTATTATAGTTGTATGTAAATGCACCCTGTAAATGTTAC
GTTGGAACCAAGAAATAGCAAGGAAATGTACAAGTTCTACCAAAGCTACCAAGTCTACCAAGTTTAACGGACTTTAACTGCTCCAAG
TTTCCATTAAATATGGGAGGGGACTCAAATTTCAGAAAAGTCTTTCACATTGGATTTAAACAAGTGGCTCTGGGTTATAAGAGTCATTCTTATATGGCACTTAAGG
GAGTTGTGCAAACTTTTCATTCCATAACAGTCTTTCACATTGGATTTAAACAAGTGGCTCTGGGTTATAAGAGTCATTCTTATATGGCACTTAAGG
AGAAAAAGATATGTTTCTCATTCTCAAAATATGCATTATAATATGCTCCTCGGGCGTTGCTCCTGGGGCGTGTCTAGAACACAAGGTATGTCTAGAGTTAAGCTCTACATTTAAGAAAGTACTTTTAAGAAGACAATTCC
CTTTAAAAATGTGATGGCTCAGTGCTCAGTACCATTTTGGAAAAGCATGAACCTCTAGAACACAAGGTATGTCTAGAGTTAAGCTCTACATTTAAGAAAGTGTTAATTGTGTGG
AATGAAAAAAATACATTATTTTGGAAAAGCATGAACCTCTTAAATTTGTTTATGCATTTCCTTAAAAATTAATTGTGACTCCAGGATACAAACCATAGTAGGCAATACAAT
CAGCTCAAAGATAAACAGTGATCACATTCCATTGCACACAGTTTGTTTATGCATTTCCTTAAAAATTAATTGTGACTCCAGGATACAAACCATAGTAGGCAATACAAT
TTTAGAATGTAATATATAGAGGTATATTTCCATATGCCCTTTGCAATAACTAGATTGTGAAAAGCTAACAAGTGTTGTAACAATAATCCATTGTTTGAGGTGCTTGCAGTTG
ACTTTGTCCATTAACATTTATCCATTATGCCCTTTGCAATAACTAGATTGTGAAAAGCTAACAAGTGTTGTAACAATAATCCATTGTTTGAGGTGCTTGCAGTTG
TCTTAAAAATTAAAGTGTTTTGGTTTTTTTTCGTTTTTTTTCCAGACATTGC

>TLK1_H_SEQID#NA_37
GCTCGCTCCCTGCGTTCTCCCCGTTCCCTCCCCGTCCCCTCTCCCCTCCCTTCCTCCCCACCTCTTCCTCCCCCCCATCCCCTT
GACTCTCCCCCCTGGTTCTCCCAGCCTGCTCTCGCTCGCCCCTCAGCGCGAGGGCGGGTGCCGGGTGCCGGTGCCGGTGCCGGTGCCGCGCAGGG
GGGGAGTCGGGTTCCCAGAAAGTAGCTTGATGAGTGTCCAAAGTGGAAGTTTGGAGGGCGCCATCTTGGTCCCCAGCTCTCCAACCC
CGGGCTCGGGCGGGCGCCAGGTCCCTGCTGAATCACGCGCCATCGGGAGGGAGGGCCAGGGAAGGTGCAATGGTCAATGAGCTTCATAGTCTGGATCCAAGA
AGGCAAGAGTTATTGGAAGCTAGAATTTACTGGAGATTTGGAAGCTTGGAAGCTAGGGAGCACTGGAAGTTGCAGTTGCAGTTGCAGGAGAATCAGCAGGGAGAAGAAACAATCAGGGAGAAGAAAGAGA
AAGCTCTAATCACAGTTTTGGAAGCTTGGAAGCTTGGGATCTTTAAGTGACAAAGAAATCAGAGAACATCAGGGGAGAATCATCAGGGGAGAAGAAAGAGA
AAAGCAGAAAAACCAGAATGAAACAGAAGTAGTCAAAGTGTCCCAACCTCCCGGTCTTCCGGCTATTGGGAAAATGGGGTGTACAGCAGGAGGAAAAAG
TATTGGGGACGTGGCCAGTGCCACAAAATTAGCGACTATTTTGAATACCAGGGTGAAAATGGCTAAGAGGCATACCCTCCTGCAATCGTCTCCTCA

AAATTCACATTCCACTCCTTCCTCATCGTTGCGACCGAATAGCCCTTCCTCCTACTGCATTAGCATTGGGGACCACCCTATTGTACAACCAAAGCAA
TTATCCTTTAAAATTATTCAGACTGATCTCACAATGCTGAAATTAGCAGCATTAGAAGTAATAAAATCCAGGACCTGAAAAGAAGGAGGACGTATAGA
TGATTTGCTCAGGGCTAACTGTGATCTCAGACGGCAAATAGATGAACAACAAAATTACTTGAAAAATACAAAGAACGATTAAATAAGTGCATATCAATGA
GCAAGAAACTTCTTATTGAAAAGAGTACAAGAAAAGCTGTCAAGCAGAGAAAGCTGTTTGCATTTCAGAATCTTGTGAAGCAACAAGAGATATTGAAAG
CATGGCGCTTCATTTACTGACGGATACAGAGATGGTTTTGCCCACAGCTAATAATTCTCAGGCAGCTTACCACAACTGTTGGATTAACTTACCACAGTCA
GCAAAGGAAACTTCTAGCAACGCAAACGCAAACCTCCACACAATTTACCACAACTGTTACCACAACTGTTGAACCAAAAACAAAGCAGTCA
ATGGAGCAGGAATGAATGAGAGGCCAGAAATCCCTTGTTGTAGACCAAAATTTACCACAACTGTTTGGCAGAATATCATGAACAGGAGAAAATTTTCAAACTTAGACTAGGAC
AATCTCAAAAGGAAGGAGAGGCCAGAAATCCCAACAATTAAATGAAGCAACTTGGAAGACCTGAGCTGAAAAGAATAAACAATGAAGA

>SGK071_H_SEQID#NA_38
GCGGTCCCTCCCCGGCCCCGGCGGAAACGCGTCCCTTTTAAGGGGGCGGGGGCTCGGGGCTCGGGGGCGGGGAGGGCCAGCGGGCCTGAGTGCCTCGAGGGC
GCCGTTCGGGCGGGGGAGGAGGATCCCGCCGCGGTCCCGACCCACCGCGGTCCCAGGGGTGGGGCCAGCGGGCCTGCGGCGCGGTCGGGTCGCTGATCATGCTTGG

GCCAGGGTCCAATGCAGGCGCCCACGCGCAGGGGGGAGCGAGGCCCAGGTCCCCGGAGAGAGCCCATGGAGAGAAGTACCAGGTTTTGTACCAGCTGAATCCTG
GGGCCTTGGGGGTGAACCTGGTGGTGGGAGGAAATGGAAACCAAAGTCAAGCATGTGATAAAGCAGGTGGAATGCAGGTGAATGATGGACCATTACGCCAGTCAGGC
CCTGGAGGAGCTGATGCCACTGCTCAATGAGCTCAGTTCAGGTGGAGATCTCTGTACCAGGATCTCTGTATCTGAGTGGATGCAGAAATGTGC
CTGCCTGGTGGTGATGGAGTTCAATGAGCTCTGGAATACCTGCAGATTTGACACATGCTCAGTGCGCTGGAATACTGCAACATCATCATCCTCATCAGCAGTGACCACTGCA
TGGGCCAGGTGCTCGGACGCGCTGGAATACCTGCAATTTGACACAGACAAAGCCAAATGAATATTCGTGCGGAGGAAGACCCCTTTCGTAAGTCTCTGGATGGCCCCTGAA
AACTGCAGGACCTGAGTTCCAATGTGCTAATGCAGAAATCAGACACATCTGTCCCTGGCTGCATCATTCTGACATGGAGGAAGACAGCAGTCGTCCTTCATGGACAGAAGCCATG
GCCCTCAACTTCTCCTTCAGCCAGAAATCAGACATCTGTCCCTGGCTGCATCATTCTGACATGGAGGAAGACAGCAGTCGTCCTTCATGGACAGAAGCCATG
CATCTGCGGAAGTCCTCCGCCAGAGCCCAGGCCGTGAAGGCCGTCCTGAAGGCCGTCCTGAAGGAGCAGATCCCGGATGTGGAAACCTTCAGGAATCT
TCTGCCCTTGATGCTCCAGATCGACCCCTCGGATGCATCACCGACATCAACGACATGCACATCAACGACATGCACATCAACGACATGCACATCAACGACATGGGAT
CTGACCCTGCACCGGCAGATGTGTCCATGCCTGCGTCCATCACCGACATGCACATCAACGACATGCACATGGCCAGCATTTAGTGATGCTGATGCCCCTGGCCCTGGCC
GGGTGCCCTGAAGCTCCTGCCATGGCCTTGGCCATGCCTGGCCATGGCCATGCCTGGCCATGGCCATGGCCATGGCCATGGCCATGGCCATGGCCATGGCCATGGCC
CTCAGCTGTGACCAGGACAGAGTCCCTGGGAAGAGACTTTGCCTCCCTGGGAAGAGGCTGTTGGGCCCATCCCAAAGGGTCTGCGTGGCC
CCCGGAGCTGGTGGAGGTGGTGTCACGGAGGTCCGGAGAAGTTCCCGGACCTCATCAGCCAGGTGTTGCCCACCTACCCTGCGTGCTGCTGCTGTGCCAGGAGCCAGCT
AGGTATCATTGTGAACAAGGCCCCTGGAGCTGGTGGGCCCCTGGACCCGGGCTGCTGTGAACAAGTGGTGGCGGCCTGTGTGCCAGGAGACA
GCGGAGTCTTGCGGTGCTGTCCCTGCGTGGCCATGGCTGTCCTGGACTGGCCATGCCTGGCCATGGCCATGCCTGGCGGCGGCAG
GAGCCCTGCTGGTGAACAATGCCTACCGGGGACTGGCCAGCTGGGCGCCGGTGTCGCTGGTGTGGTGTCCACCTGGCTTCCTATGAGG
TGGCCTCAGCCTCATCAGCAGAGAGACCTACCAGCTCCAGGACTCCACAGGTGGCCCTGCTCCAGCCGCTGCCACTGGTTTCTTGTGCGGAAAAA
AGATCCTGGAGCTGGTGTCCAGTAGTGAAGGCCCTGCTCCAGGAGATCAAGGGAGCTCCAGGGGAGGAAGACAGATACCCCACAGGCCCCTC
GTGTCCACGTGTGCCCTCCTGCCCTTCCTTTCCATGGGCCACTGTTCCTTGCCTTAAAAAAAAA
TCTTAGGTGACTAATAAAGAGGCCCAAGGCCAGTTTCTGCCTTAAAAAAAAA

>SK516_H_SEQID#NA_39
GCGGGCGCACTGCGGTTTGGTTGTTTGTCAACGGCAGTGACGGAGGTTGGGAGCCAGTCTGACTGCAGGCGGGGGCGTCAGGAGGAGGCGGATTCCTCCTCCCTT
CGGCACTGGGGGGGGCGGAGGCCCTGCGCGCGGGCCGGCCGCGGGCGGGGAGGGGCAGAGAGCAGAGGCCGCAAAGACAGAGGCGATGGAGGGCGACGGGGTGCCA
TGGGGCAGCGACGCCCGTCTCGGGTCCCGGCGGAGGCCCAGAGTTCTTCCGGCGAGAATGATCCGGCGGAGTGTGCGGAGCTTCGGCCGCTACCTGGGACGGGCT
GCGACAGAGAACCTGCGCGAGGACACCGGAAGTTCTTCGCGGCGACATCAAGTGCTCCCACCACTTGTCTCTCTCCCCACTAAGGAGAAGAAGTACCTCCAGCAGATTGTG
GCGGGCCCTGCCTGCAGGCGGATGTCGCCGAAACGGGCTGCAGGCGGGCCAGAGATTGTAAAGTCAAGTGCCAGTCTCAAGTGTCGTGTTGGGGGTGCAGGGTGCTCCCACCACCAAGCTG
GACTGCCCCTTGCCATACTGCATCTCGCGGCCAGGATTGTAAAGTCAAGTGCCAGTCTTCACCTATGGAACCAGTATGAACTAGTGCACAC
GGCAGTGAGGAGAGCTGTAAGCTGCAAGTGGCAACTGGGACCATGCGAAAGGAGGAGGATCTGGAGGTCCAAAGAACAATCGGGCCACAGTGGAACTGGAGG
GCTGGTTGCTCATCAGGGCACATGCTCTTTACAGGAAGTGGACGTTTGTGGTAGCACGTTGTGGTGAGGATGTGTCATGTTTAGCGGAACTGGTGAATGATT
TAACGGATGCACCATGCTCTCTTACACTCCACAAAGATAAACTCTCTCCAACCAGGAGGCTTCAGGAGAGCGCAAGAGATCACCGCTTTATCTCCTTTCCTGTATTCTT
TTTCAAAAGTGCCGAAACTGGGCTCGGAGATAATAGACTCTCCAGGAGATACTAAAGCTCAACAGCATGTTGGGAACAGAGTGAAAAAGCTGAGCACGTGGGCT
ATCTGAGCAGCAGCAGCAATGGGGCGGCTCCTGAATGTCGCCGAAACCGGGCCAACTGGAGCTGAACCGGGCCAACTGAGTGAAAAGCTGAGCACACA
TTTTCTCACCAGGTGTTACAGACTGCACTCACCCCAAACCGTGGGGCTCCTGATGCAGCCAGGCCCTGAACTCTTTATTAACCAGGCATTTGACA
TGCAGCGGGACCTGCAGATCAGCATCCTCCCAAACGTCTGAGAAAAAAGGAGAACTTCTGACAATGTGATGATGACTACTAACAACTGCAAAGCAG
GAGGAAATGAAGGATATGATGATTGTTGAGACACTTAATTAACAACATGGAGGTTGTATGAATATGCCAAGACTGTCATTTGCCCTGA
GAATGGAGAACCAGTGACGCACCAGAGAGTCAAATGCTGCATCAAATGCTGTCAGGCAAGCAGGACAAGCTCCCGACTTAATCAGGCAGTGGCTAATAAGCTGA

```
TCAGCTCAGTGGATTACCTGAGGGAAAGCTTCGTCGTCGGAACCCTGGAACCCTGGAAGAGTCTCAGGATGTCTCAGTTCACATCACCAGTA
ATTATCTCAAACAGATCTTAAATGCTGCCTATCATGTTGAAGTCACGTTTCACTCAGGGTCGTCAGTTACAAGGATGCTATGGGAGCAAATCAAACAGATCA
TCCAGCGCATCACATGGGTGAGCCAACCTGCCATCACTCTGGAATGGAAGGTGGCCCAGGAAGGCCTCAGCGCCTCCAAATTGGCT
AAGAGCAATTGCAGCCAATTCCGGACTCGGCTCAATAGTTCCACGAGCGAGCTTTGCGGCCACTCGGCCACTCAGGCCGGTTAGAG
AAAACGGAAGATCTATGGCTGAGGGTTCGGAAAGATCATGCTCCCCGCCTCCCGCGCCTTCTGTGACAACTGGGGAGGACACTTTCCTGTGCCCTCAAATCAGTTGTC
AAACCTAAACTGGGACAGGAAGCACTGGAATGATCTGGCTTTGGAATTCACTGAAGTGGTCTGCGAAGCGATGGTGGGATCTCCATGTTCAGTCATT
CCTCCAGATGGAGAAGCACTGGAATGATCTGGCTTTGGAATTCACTGAAGTGGTCTGCGAAGCGATGGTGGGATCTCCATGTTCAGTCATT
GACTACAACTATGGTGGTGGCTCCAGCATTGCTGTCTGCTGCAGCATTCCGCTTCCTGCACAGGGACTTGTCCATCGTGATATCAAACTGAAAAATGTGCTGTGGAT
ACACGTTTGCAGATAGCACTGAGATGTGTGGAGGAGAAT CCGGTCTTCCTGCACAGGCCATGATGTGGGACCAATCCATATGGCCCTGAACT
AAGCAGAAACCGTGCCAAGTACGATACGATAATTCCGTGGATGTTCGTGGATGTTCTTTTGTGTATAATCTGCTCAGCCTCGTCAGGCTCGTCAGGCTCTGTCAGTGAGGCATTGAGAGG
TTTCACAGGGGAAGTACGATAATTCCGTGGATGTTCGTGGATGTTCTTTTGTGTATATCTGCTCAGCCTCGTCAGTGATGGAAGCCTGT
TGTGCTAGCAAAGACCATCTCTGGAACAATGTGCGGAGGGGGGCCTCTCAGCCCATTGTCCAGCCCCATGCTCTGCAGTCCAATCTCGAGCAGCCAAAC
TGGGATGGCGACCCCTGAAGAGCCCTCTCACTTGAAAGCCAAGACCTTCTACTGAATCGGCTCTCAGTTGCCATGGGGGAGAATTTGACATTTAT
AGAGGACTAGATGATTCTACTTGAAAGCCAAGACCTTCTACTGAATCGGCTCTCAGTTGCCATGGGGGAGAATTTGACATTTAT
TCACTATAGGACACACTCCCAGGAACTGGTGCTGCTGGGAAACCCTTCCCCAGGCAGGGATGATGATGATGATGATGATGATGAATGTGGTGTTCTCGGCTTCACAAATG
AGCATATTCAGCGACTCACTGAAGCGCCCAGCTATCCTTTAGCTCAGTGTCTTTCACTGTATATACTTTCACTGTATATACTTCTAAGACAGCAAGCGGGAACACTGCAGTGCAATAGTGTTAAAAATCTCATTCTCATG
AAAAGGAGGCAGATGTTACCATTGTCTTTTCACTGTATATACTTCTAAGACAGCAAGCGGGAACACTGCAGTGCAATAGTGTTAAAAATCTCATTCTCATG
ATTTTTGGCTCTAGGTACTTAGTCAGGAATAGTTTCACTGTATATACTTCTAAGAGACTAGTTCTGGAGTTCAGAAAATTTGAGCTAGTTCTGGATTATGAAAACAATAACTTGTGGATG
TCTGGGTTCCTGCGTCTGCAGCCAGTTATCTCCTGGTTTCTGGGGAGGGTTCACTGTTCACTGTTCACTGTTCACTACGAGGAGAGTCTGTTCCTCATAGTGCCAGTG
TGTTTACATTTATAGGACCACATATCCCTGGTTTCTGGGGGCTTTGGGAGGCTGACACTTTGGGAGGGTTCAGGAGCAGCCTGGCCAACATGGTGA
CCAGGCATGGTGGCTCACGCTGTAATCTCAGCACTTTGGGAGGCCGAGGCGGGCGGATCAGGAGGTCAGGAGTTCGAGACCAGCCTGGGCCAACATGGTGAAC
AACCCGTCTCTACTAAAGATACAAAAATTAGCCGGGCGTGGTGGCGCGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAAC
CTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCACGCCATTGCACTCCAGCCTGGGCAACAAGAGCGAAAACTCCGTCTCAAAAAAAAAAAAAATAGAATGT
GTATGCATATATTTGTGTGTATATGTGTCTTTCTTTGTTTACCTTGAGTCTTTCTTTGTTTACCTCTATATCTCATCTCAAAGTTGGAATTAGAGTTACTGCTAACC
CCAGAACAGAATCGTTCTTTTACCTTGAGTCTTTCTTTGTTTACCTCTATATCTCATCTCAAAGTTGGAATTAGAGTTACTGCTAACC
TCTAAGCAAAACCAACCCAATGCGAAGTATTTGGGTGAGAACTGAGAGTGGGGAGATGATTGGAGGGCCATGGCCATGAAAAC
CAGCTTGTTTTAAAACTTAACACAATCAGAAAATTTTCTTTTTTACCTTTTAAAAATCTGTTAGAAGAAAAAAGCAGACGTTTGAACTGATCAATGAGCGTTTACCACGA
TTTCTCCCGCAATTGTTCCTCTTTTGTTGTTCGTGTGCTAGTGTTCCCTGAAAGTTCCCTGAAAGGTTTCCCATCTGTTTACCTTGAGAAATCTGTTAGAAGGAAA
TGCCCCTCACCCTGTGTTCTAGGTGTTGTTCGTGTGCTAGTGTTCCCTGAAAGTTCCCTGAAAGGTTCATGTTCCTTCCATTAGA
AGAAATTCTGATTCTAGGCAGAGGAGGAACAATTTAAATATTGGTGCATCTCCCTCCAACTGTTCGCCCCTGTTCAGCCC
TTCCGAAAAACAGGTAAAACAGGTAGACGAGGTAGCAGTGAGTACAGTGAAGTGAACAGTGAAGTGAGGGTCAACTGGTCGCTAATCCCAATAACTGTCGCCCTGTTCAGCCC
CAGAGTTTGTAAAACAGGTCCCAGTAGTTGTAAACAGGTAGGTTGTATAAGATGGTATAAGATGGTGTAAACAGGTTGTAAGTGTGGAGGCTG
AGGCAGGAGAATCACTTGAACCTGAGGTCATGGGTGTATAAGATGGTGTAACATCGTGTCTTTCTGGGATTTTGTAGGGTTGTCTATGCATGATGAGGAGACTCTATCAAAA
AAAAAAACAAAAAAAAAAAAAAAAAACAAAAAAACAAAAAGATGAGTCGCACATCTGCTCGTGGGATTTTGTAGGGTTGTCTATGCATGATGAGGAGACTCTATCAAAT
TAATGGAGTATAAAATGGAGTCTATAAGTCTATATAAGTCTTTCATATAATGTGTCTTTCATGGGCTCTGATCATGTATATGTTTGAGCTAGTAGCTCCCT
AACTCTTGGAAGCCTTTCAGAGAATTCCATGTAGTGCTAATAAACCTTTTTTACTTGCAAATGAAATAAGTAACTGACTCAGTGTCTTTGGGCTCTGCAGGCTTAG
CCTTGTTTTGGTAGGATTGTCAGTATATTCGAGAATTTTACCACCTATTTTACCACCTATTTTAGCTATACTTCAGTCTATACTTCTCCTCTGGGA
ACTATGATGGGGCTTCAAGGGAGTTCAAGGTGTCAAGAAAGGGTAAGTACCTCAACTCATTGTGGAAATCATTGTAAGTCTGTTGTGTATGAAATTTTAG
GAAAAGTCTATATAGTTTCCTTTCCAGACTGTTTCTTTCTTTGACTCTAACATAATCATAAGTCTGTGTTCCATTATGAAGGTCACATTCTCAGCATTCTGG
```

```
GAAGTATTAGCTCATACTCTCATTCACATATGGGAGGTTAGAGATTTTTTCCTCTACCCAGAGAGTCCCTTTTTAGTTTAGCTACCCACCGGCTTAGTTTGCTGTCAGATA
CCAGTTCTCTAGCAGATTCCCCAACAGTTTTTCCCCAGTGGGGAGGGTGGAAGGGAGGACTGATATAGTAAGTAAGGAGATGAGTAGTTGTGTTACAGTTCAT
TTTCTTTAAAAAAAGTACTTTGAGAAAAGGTGCAAAATATTCTTGAAAGTCTCTAGCCAGCCAAAGTCCACATATGATAATTTGGGATGTTATTTAAAT
ACTCAGCAGGCTGAGTGCAGTTGCAGGTATACGTGTGTATGTATGTATATTACTTGCTTTGTTTTATAACATTGTCAGCTAATTTGGCTTTGTTCTTTGGCATCCAT
TCTTATTATAAGTGATGGAATGCCATGAAATGTACTGTATTGTATATTACTTGGGAGGACACTAGGTTTCAAAGAAGTCTCGGCAGCACTTGGGTTATAGAG
GGCTCAGCATGACCAAAGTCCTAAAGATTTTACCTCCCAGCTTGGTTGTATGTGTCCCAAGTGGTGAGTTGGCTGGAGGAAGAAAAAGGATG
GCATGAAAGAATGACCCTTATTGTGTCTGTCTCAATCAACATAATTAGGGCAGCTTACCCAACATGAAAATGATGACCTGGCCATTTTAATGGCCA
GAGCTTGAAGTCTCGGTCTCTCATTTGTTTCAGAAATGAGTGAGAGGCCTTCCTGCAGAGCCAGTCAGGAGAAATGGTGGAGCCTGGAATGGCTG
ACCCTGTGCCTAGAGAGCTTCAGGAGGACTGTGCTGGGGTTCCCCAGGCATACCTTAGTTCCCAGCATCCTCTGCAAAGAGTA
GCTCATTGACTGAGCAAGCGCTTCAGGAGGACTCACTTTGACTCCATCAACTTCAAGGGGACAGAGCCTAGG
CAGCAGATTCTCCTAATTTTTAACACTAAAATTCAAGGAAGCAGTGGAGGAAAAGAAATGCATTACTACCACAAAGTGTTCTTCCTCATGAGAAATAGGGGACTCATAGAACTCTCG
ACTAGTTACCTGTCAGCTCTCAGATGCATTCTTGTGTCTTTGTGTCTGCCAAAGAAGGTCACAACTTGGGAGGCCCAGGGGAATTGGTGATTACAGTAGC
CAGGGCTGCAAAGTGCAAACTGCACTGAACAACACCCCTTTGGGCTTGCTTGGAGAGATGATGGCATGCTTGGTGTATTCCATTTCTTGTATTCCATCCATATT
ACTTTGGGAAGTTAAAGCCACTTGCTTTTGGGAGATTAACAACCCTGCAAAGAGTAGTTGCTTTCCAAAGCTTTTGATTTTCTGATCAAGTAACCC
GTGTCAGAGTGGGATTAACAACCTGCAAAGAGTCCACTGCCAAAGAGTAGTTGCTTTCCAAAGCTTTTGATTTTCTGATCAAGTAACCC
ACTTTTTAAAACCTGAGGTCAGGAGTTCAGGAGTTCGAGACAGCAGCCTGTGAACCCGTCTCTATTAAAATAGAAAAAATTAGCCAGGCATGGTGGGG
GCACCTATAGTCCCAGCTACTCAGGAGGCTGAGGTAGGAGGAATTGCTTGAACCCAGGAGACAGAGGCTGCATTGAGCCAAGATCACCACCACTGCACTCCAG
CCTGGGCAGCAGCAAGACTCCATCTC

>H85389_H_SEQ1D#NA_40
CATGACAAGCTAACAGAGAGAGAGAGAGAGTCACAATCCAAGGTCCGGCTCCTCCGCGGTCCCAGGGGCCGGACGGAGGGGATGAGGCAGGAGGGGGCCCGGGCA
GCGCCGGTTGCTGCTCCCCCCGGGGGGCCGCAGCCAGCCGGCAAAGGGGAAGGACGGGCCGCAGAGGTACACAAAGCCCGGGAAAAAGGCGAAAGGCGAAGCCCAG
TGCCGCGGGGGGCGCCCAGCACGAAGCTGAGGGCGTCGGCGGCCCATGGCGGCGCGGGCGGCCCTTCCTGGGCGCGGCG
ACGGGCCTGAAGCGGTTGCCCGGGCTGGTGCCCGGCCTGTTACAGCGGTTTATGAGGCGCGGTTGTTTATGGAGGCCAGTGCCGGGGGC
GCAGCGGGGGCCCGGGTGCGGTGCGGTCAAGAAGATCCGGTCGTCGGCGACGGTGGAGCGTGGGCGCTCGGCGAATTCGGCCCTCACCAGCCTCAAGCGG
CGCCACCAGAACGTCGTGCGCTCAGTTTGAGCGTCCTGCAGTTTGAGGGAAGAAAGGATCCGGTCTTAGCCTGGCGCTCGCAGCTTCTGTGAAGGTGGAGACCTGAA
TCAGTATGTCCGTGTCCGGAGGGCCAGGCAGACAACATCCTCAGTCAACAAAGTTCATGCTGCCACCCCATCCAAAGTGGCCGGAGCCTTTGGACTAAGCAAGGTCTGTGCTGGGCTGG
CAGGGACCTGAAGCCAAGAGAGGGCAATCAAGCAGCCAAGGCGGCAATCAGCAGGGACTGAACAAAAATGTGAATAAGTGAATAGAAAACCATATTGTGCA
CACCCCGGAGGGAGGGACACTACACAGCAGGCGGACTACATTAAACAGGGGACTCAAGGCTCTAACGGGATCGTCCCTGTTGGTGAGGCGTTCGGACTTCTACAGCTCCTGAA
AAGGAGCTCCTGGGACCTTCCATGTCGAGGGGATCAAGCAGCTCTTGAAAGATATTTGGGTGCTAACTGAGTTGCACATCCCCAAAA
ACGCAGGACTTCACGTCTGCAGTCTGCTCTTGAAAGATATTTGGGTGATTTAAACCAGGCTGATCCCGACAGTCGGCAACCACACGTCTCCTCCA
GGACCAGTGCAGAGGGTACAGGTGGTGGCCGGTTGGCCTGGCCGGTTGGGCGATCCCGACAGTCGGCAATGTGAAGCTTTTGTTTGGGTTTCCCGGCTTCTTTT
TAGTTTTGCTTATTTtttCCTTTTGTTTttCCTTTTTTTTTTTAAACCATTGAAACTTCAGAAGAGCAGGACACAAATGCTGTGGACA
GGCACCAATTTCTTTAAAGAAATTCAAATGTGGGCAAGGCATATGGTAAATTTCACTTTTATAAGGGGTTAGGGGAGCTATTTTGTTTGTCCTTC
ACTTTCCCTCGTGTCTTGTTCCTTCTTCTTTATACTTTTCTCACTTTCTACACCTCACTTCATGACACCTCCCTAGAGAAGGCCTGCCTCCCATAGGGAATCTGGGGGTTTCTTC
```

TGGAACGGGGGCGTGAGGACACAAGGAGGAGGCCTCTGGGCCACGCCTCCCTACCAGATGCAGGAACTCCTGGACTCCTTGGTGGGCTGGCCCTGCGTAGCCCTT
GGGCCCTCGGAGATGATCAGAGGTGAAGAACCGCC

>Weelb_H_SEQID#NA_41
ATGGATGACAAAGATATTGACAAAGAACTAAGGCAGAAATTAAACTTTTCCTATTGTGAGGAGACTGAGATTGAAGGGCAGAAGAAGAAAGTAGAAGAAAAGCA
GGGGAGGCTTCGAGCCAAACCCCAGAGAAGGGTGAAGTGCAGGATTCAGAGGCAAAGGGTACACCCTGCAAATGTCCTGAGACACCAACCAGACCAGAGCA
CACATCTTCGGAAAAAGACAAAGAACAAAGAAAGTCCAGATTTTGAGGACTCCAGTTCACACCCTCAAATGTCCAAACAGGGAAGCTTCCTTCCAGAGGCCCTAAGC
GGAGCAAGCTGCTGCCCAGTGACAGCAGCGCCCCTCTACTCCAAAACCATGCTGAGCGCGGTTGGTGCAATATTAATCCCTCACTCAGAGTCCTATAAAAATTATTTCTTCA
ATTTGAAGCTCACACCTGCTCCCCTCAAGGATGACCTCATTGGCTCTGGTCAATATTAATCCCTCACTCAGAGTCCTATAAAAATTATTTCTTCA
ATCTGGTGGCAAGAGAGAAAATAAGAGGAGATCTTGAGGAGTTGGCTGTCCAGAGAAGGCAAGTGGTCCAAGAGATGTGTTTTACGAGAAACC
AACATGGCTTCCCGCTATGAAAAGACTCTATGAAACTTTTACAGAATTATCAAATGAGAATTCGGCTTTGTCCATGAAGTCTCACGCAGTGCTTGGGCATCACCCC
TGTTTATGCAATAAGCGCTTCTATTCCTCACATGAGAGACAAACTTCAAGATGCAAATGCCACATGATCCTTCAATTACAGATTTCCCTGCCTATAGAAGAAGTTGCAACTGCACCTGGT
ACTAAGTCTGGCAATCATTTGAAGAGCCAAAAGACCATAGTCAAGATGCAAACATCAATAAACAAACCAAAGTGAAGAAGAGATAGTGCTTCCTGGCTAATGAGATTTTGCAA
ACATCAAACCTAGTATATTAAAATTGGTGACCTGGCCAAAGACAGAGCATTAAAATTGCCTTGGGATTAACAATTAACAGAAAAGCTTTCAGAAAGCTTTCCCAGTCAAGCAGAGAGA
ATGTGATGTATAAAAATTGGTGACCTGGCCAAAGACAGAGCATATTTGCCTTGGGATTAACAATTAACAGAAAAGCTTTCAGAAAGCTTTCCCAGTCAAGCAGAGAGA
GAGGATTACCGGCACCTTCCAAAGCTAACTTTCCGGAGCTCTCAGAGGCTTCTCCGGCCTTCTCCGGAAATATACAGTTCTCGGGAAAACAGAGGATATACCTCAGGGTCTGGGGTCTGGGGGTCTTGGGGACCCACACAGGATCA
CACCATATCGCAGCAGTCTGGCCAGAAATACAGTTCTCGGGAAAACAGAGGATATACCTCAGGGTCTGGGGTCTGGGGGTCTTGGGGACCCACACAGGATCA
AGAAGCACAAAACGCCTGGTGGGAGGAAAGACCTGGTGGGAGGAAAGACGCCTGGTGGGAGGAAAGACGCTTTACCTCAGGGAGCGTTACCTCAGGGAGCTTTACCTCAGGGAGCGTGAGCCTCGCATTAA

>Wnk2_H_SEQID#NA_42
CACGGGTAAGCTTGGGCCCCTCGAGGGATCCTCTAGAGCGGGCCGCCGAGACGTCCCGCCGCACGCTGATCGAGCCGTGATGAGCCGCGGGGCCCGCGGGCAT
GGCGGAGCCTCGGGCGAAGGCGGCGGGCCGCCCCGCGCGGGCCCAGCCGCTTCTGCGGCGGGCCGCCAGCGTGCTGTAGAGTCGGACCAGGAGGACGGCCGGCGCCCGGGCTTGGAGGCGCA
GCCGAGGCGGCGGGCCCCGCGCAGCGGGCCCCTCATCGCAAGACGCGCCGCCTCATCGCGGAGCGCCGGCCCGGGCTTGGAGGCGCA
CGGCGCCCCGGCGCGGGCCCAGCGCGGTCGGTAGCGCAGCGGCGGGCCCCGACGGGGCCGCGGTGGCCAGGAGGAGGATGAGGGGGGCGGCCG
GACCCCATGCGAGCCGAGCTGTGTGAAAACCGCGGCGGCCACTCGCTGCGCCGGGCCGTGAGGAGGCCGTGGCCCACCTCTGACGGCGCCG
AGGCGGAAGCCGAGCTGACATCGGCACGGCTGGGCTGCGCGGGCCGCGGGCCCGCCGGGTTCAAGGAAGAGGCTGAGGGCTGAGATGTCGTGGGCGCCACCTCGCGCTTCTACGACTTC
CTTCCTCAAGTTCGGCGACACTCGGGAGCTGGGCGTCAATGCGGTGCATTGTGCTGCGCCAGATCCTGAAGGCGTCGGTGCCGAGGCCGGTTCAAGGTGGTTCAAGGTGATGA
AGCCCAAGGTTCTCCGCACCGGAGCCAACTGGGTCGGTGAGGAGCATTGGCCTGCGGAGATTGGCCGGACTTGGCGCCAGATCCTGAAGGCGTCGGTGCCGAGGCGGTTCAAGGTGATGA
AATATTTCATCACCGGCGCCCGAGATGTACGAGGAGCACTACGCCCAGAGATCTACCACGGAAAGGAGGAAGTGCCACCTCGGAGTACCCC
GAGTTCATGGCGCCGAGATGTACGAGGAGCACTACGCCCAGAGATCTACCACGGAAAGGAGGAAGTGCCACCTCGGAGTACCC
CTACTCGGAGTGCAGGCCGCCAGATCTACCGCCCAGATCTACCGCCCAGCTTTGAGCGCGCCAGCTTTGAGAAATCAAGGATGA
TTATTGGGGAGTTATCTGCAAAAACAGGAGAAAAGGTTACGAGATCAAAGACCTGCTGAAGAACTGAAGGAAAGACCGGCATGTGGAGGGGTGA
GCTCGCGGAGGAGGACCACGGCGCGGGGATGGCCCAAGGTGGGTGGGTGGGTGGTGGGGTGGGAAGGACCATGTGGAGGGGTGA
ATAGAGTTCACCTTCGACCTGGAGAGATAGAGGTGGCCCAAGAGATGATTGAGTTCGGAATCGTTTCCACGAGAGTGACGTCAAGATCGTGGC

CAAGTCCATCCGTGACCGCGTGGCCTTGATCCAGTGGCGGCGGGAGAGGATCTGGCCGCGCTGCAGCCCAAGGAGGAGCAGCAGGATGTGGGCAGCCCGGAC
AAGGCCAGGGGTCCGCCGGTGCCCGGTGCCCCTGCCGTTCCCTGCAGGTGACCTACCATGCACAGCGGCTACCAGAGCCCGAGGAGCCGGAGGCCGACC
AGCACCTCCTGCCACCTACGTTGCCGACCAGCGGCCACCTCCTGGCCTCGGACAGCAGCACCTTCGACAGCGGCCAGTGTGTGCAGGGCCGTCCTGCCGCGC
GCAGCCAGCAGAGCGTGATGCTGGGGGCCTACCAGCAGCGCCCCGTGCGCGGTTGCGGGCTGCTTGGCTCCACGCACCCCCATGCGCGGCCCCGTGCTGCGCG
AGAGCCTGCCCTGCGGGGATCCCGGATCCGGGGATGAGCTTCGCGCGCGGGCCCCACAGCCGTCTGCGCCCTCCAGCCGTCCCTCGCCGCCCGGCGGCCAGCAGCCTC
ACTTCCCGGATCCGCCAGCCATGAGCTTCGCGCGCGTGCTGGCCCCCGTGCTGCGCGCAGCCGTCCTGCCCCCACGCCGGTTCCCCTCCAGCGGTTACCTGCTCCAG
CTCCGCTGCGGGCCAGCGACACCCCCGGGCCGGCCCCCACAGCCGGTTCCCCTCCAGCCGCCGGTGGCCCCCTGCTGCCGCCGTTCCCGCCGTACCTGCTCCAG
CCTCCAGGTGGGGCCCCCGTCCGGCCCTCGGGGAATCAGCGCGCGCGCCCTCCCCCCCTGCGGCCCCCCTGCGGCCCCCGAGGATGCCCCCACCACCTCAGTATTT
TTGTGCCCCCATCACGCCCCCTGCCGAGCGTTGCCGCTGCCCTCCCCTCCTGCCCGGCCGCCGTGCGGGGCCCCCACCCCCCTGGGGCGCCC
CTCTCCAGCCGTGATCTTGCGCGCGCGGCCTTGCCGCCGCCCTCCTGCCCCGTCCCCTGCCAGGGCCGTGGCGCCCCTGCCGTCCATTCTGCGCG
CTGGCCCATGCCCTGCCCGGACCATTGTGCCCAAATGCCACCGGCCTTCCCCAGCGGTTCCCCCATCACCCGGGTGCAGAATATGAGGG
TGGCCCAGCTCCCCAGGCCAACCTGGTTACCCGAGGTCTCCACGCCCCCAACCACGCTGCCCCCTCAAACCGTGTGTTGCCCCGC
CCACCCCTCCACAGCGCGGCACTGCCTCCACAACCTGCCCCACAACTGCCCCCAACCTGTCGTGGCCCCGCCACTGCCGTGTGTGCCCCTGAGCCCCC
AACCACACGGGCCCCTCAACCTGTCTGCCCCCGCAACCTGTCGTTGGCCACCAGCCAGATTCTGCTTGGCCACCACCTCCTATGCTCCGTGTGGACGTCGCCGCTCAGG
TCCAGCCCCACTTCTGAACAAGCTGCTCACACACAGGGAGCCAGATTCTGCTTGGCCACCAGCTCCTATGCTCCGTGTGGACGTCGCCGCTCAGG
TCCCCACGTGCCCTGTCGCCACCGGTCCTTGCCGTGCCTCTGCCGGAAGTGCTCTGCGCGGCCCTGAGCTCCTCAGTTCCCCAGCTCCCT
GGCCACGGTGTCTCGTGTGCAGAGTGTGTGCCCACCGCCCCCTGCTGCCCAGCAAACCCACCGGCGCTCCCGGCGGGCCCGGATCGCCAGCCC
TTTGCCCAACTGTCCAGCTGACGGTGGAACCAGTCAAGAGGAGCAGGCCTCAAGGAGGTGCAGCTGTGAGAGCTGTGAGAGCTATGGAGGTT
CTGATGTCACTTCTGAAAAGAGCTGAGTGACAGCTGACAGCTGTGAAAGGGCCTTTGAGGGGACGCTGTGAGAGCTGTGAGAGCTATGGAGGTT
CACGCGTGCGGCTCCCGGCAGGAGAGGGCCAGGAGAGGGCCAGGTGCTTACCATCTTGAAACGTATGTGCAACACTGGGTGGAGTGCCAGCTGGAGA
CGCACAACCACAAGATGGTTCAGCTGGACCTGAGAGGGCCAGCCCTTACCATGGAACAAGCAGATGGTCAGACGCCGTATCCTGCAGGCC
GAGCGGGAAACGTTCATGGAACCTCAGCACCTGCGGCGTGCTGGCACCGGGGGAGAGTCATGGACCAAGGCAGAATGCACAGACCGTGCTCCGACCCAGGGA
CCAGCCCGCCACACCTCAGCACCTCAGCTGCCGGGTGGCTGAGCACCCGAGCCCCGAGGCCGACAATCCCAACGTCCTGCCGCCAGAAGCC
GGGAAGAGGTGGTTCATCATCGTCCGGGTGCTGAGCAGCTGGCTGAGGAACAACCCAGCTTTCTCTCTAAGCTCCTGCCGCCGGGCGGGCCCCAGCAACCC
AGCCAAGATTCAGCGCCGCTATAAAGAGACCAGCTCCCGCGCCTTTCGGCGCCCCTCTCTGAGGCTCCAGCAGGTCTCAGCCACCATGCCAGAGCCAGC
TCCTGGGGCACCCGCAGCGGGCCCTTTGGCCCAGGGCCCCTCAGTGGGTGGCCTGTCCCAGGAGGGTCTCAACCAGGAGGGCCCAAGGGAGCCCTGCCACCT
GTCAGGAACTGCCAGCCAGGCAGGGGGGTCACGGGGACACCTTGCACTCAGCTCCAGGAGGCTGCCTCAACCAGGAGGGTCTCAACCAGGAGGGCCCAAGGGAGCCCTGCCACCT
CTTGCTGTGCAGCCCCTCGTGGTGGCCCTAGCACGGCGGACCCCACACCTTGCACTCAGCTCCAGGAGGCTGCCTCAACCAGGAGGGTCTCAGCGCGGCCACC
CCTGCACCTGAGCCTCAGCCCCCCACAGCGGGCCCCACGCCCCCCACAGCCCCCGTCCCCCAGGAGGCCGCCCCAGGGGTCAGCGCTGGCCACC
TCCCAGCTCCCCAAGCCCAAGCCCACCGCTCCCCCCAGCCCCGTCCCCCAGCCCCTGGAGTCGCCCACCTCCTCCACCAGGACACGTGGGAAGGGCGCCCCGTGGACCCCTGTGAGGTGGGCTTTGTG
GACAGCACCACCATCAAGAGACCTGGAGCAGCTGCTGGGGAGAGAAGAGCTGCCCGCGTGCTGCTCTGCCACCTCCTCACCAGGAGCACGGTGCCCACCTCCTCCACCAGGAGCACGTGGGGTGGG
CGACAGAGACTTCACCCTGAGCCCGTGCTGGGGGACCAGGTCTGCGGGGGACCTGCGTCCCCAGGTTGCCTGCCTAGGAGGCGGTCT
CAGGGGCGTGTCCAGTCGTCCAGCTGCCCTTGGTGAGAAGTCAGAACTGCAGAGGTGCCCCACTGGCCCCAGAGGCGGGCCTCAATGACAGAG
TCGTTCTCCCAGGACTATGCTAGGCTATAGTACAGAGGCAGGGGTGGTGACACTCGGTGGTCCCCAGGATGTACCTGCTTTGTG
AGAACCTGCAGTGTGGAGCCCACAGACCCACAGATGGTGGAGAAAGCTGGAGAAGCTGGAGAAGTCGGACAGAGAGAGTCGGACAGTGGGGGCCAGG
CTAGCCACCCCAGAACTCCAGGACCTCGGGCTCGAGCTTTGGGGTCCCCCACAGACGTCGACGATGTCAGTCCACCCGAGGCGTTTCT
CGGTGGTCAGCACTCAGGACGAGTGGCGGCGGGCCAGAACGGCGCCGGTGTCCAGCACTCCGAGTCGGGGACTCTGGGGACAGGGGCCCTCAGCCCG
ACGTGAAGCTGGCAGTGGCGGGGGCCGTCCCTGCCGCAGAGCAGGCAGGCCAGGCCACCGCCGTGTCGCAGACTCTGGGGGACAGGGGCCCTCGGGGCG
AGACCCCGGTGCAGAAGCAGCAGGAGCGTCCTGCGCCGTGGCGACTTCGTTGCGCGACTTCGTTGAAGAAGGCCACCGCCGCCTTCTCGGGGGC

CGGCTCGCTGGGCCCGAGACACCCAGCAGGGTGGGCATGAAGGTCCCACGGATCAGCGTGACCTCCTTCCAGTCGTCCTACATCAGCAGCGACA
ATGATTCGGAGCTGGAGGATGCTGACATAAAGAGGAGAGCTGCAGGAGTCTGCGGGGAGAGACGCCACCTGGAGGAGATCTCGGAGCTGCAGAAGCA
GGAGATGCAAGCTCTGTACCGGCCGCTGGCCAAGCCACTGCCCCCAACGTGGGCCTTCTTCCACACGCGCCACCCCCACTCACTGGCTGACCCAGCCCCCT
AGAGCAAGCTGAAGCTGAAGGCAGGCAAGCTGCTAAATCCCGTGCGGCAGCTCAAGGTCGTGCGGCTGACTGAGCGGGACCTGCTCGTCCGTCCGTCGTGGGC
CCCGCTAAGGACCCTGCCCAAGCCCAGCCAGTGTGGGGCTCACTGCAGACAGCAGCCAGTGCAGGGGGGGGAGGCCGTGGCCAAGGCTGCAGACGGTTACCACCCAACGTCTGGCAAAGAACAACACAGCAG
CTCCCTGTCTTCGGACATCTGCTCGCCCGGTTAGCCAGTGATGGAGGCGGAGGCGTGGCCAAGGCTGCAGACGGTTACCACCCAAGCGTCTGGCAAAGAACAACACAGCAG
TAAGTCTAGTAGCAAACCTGGTGCTCGATTCCTCAGTGGACCCGTATCTGTGTCATCTGGTTCAGTGCCGTGTGGGGTCTGCCCTGAAAGCGTCTCTGCCTGGAGGCCAG
TAGGTCCTCCACCAGCAGCCTGGCCCTGAGCCAGTGCCCTGCACGTCCCAGCCCTGCAGGGCCGCAGCGACAGCGCTGGTGGACCCCAGCAAGCTGTGGACGAAGAGTGGGGTA
CCTTCACGGACGACCTGCACAAGCTGGTGGACACAGTGGGGCTGCCCTGGGCTGCCCTGGCGGCGGGGGCTATGAGCCGCTCACGGAGTGGGGATGCCCACCGTCTGCCC
GAAGCTGCAAGACATGGAGAGGCCAGGCAGGCTGGGGCTGCCCTGGGCTGCCCTGGCGGCGGGGGCTATGAGCCGGATGGGGGATGCCCACCGTCTGCCC
CCAGCGCCCGGCCCTCTTGTTCCGAGTCCCCCCACCGGTCATTCCCGAGTCCCCCACCGGTCCCCCCACCGGTCCCCCCACCGGTCCCCCCACCGGTCCCCCCACCGGTCCCCCC

>MAP3K1_H_SEQID#NA_43
ATGGCGGGCGGCCCGGGGAATCGCGGCCTCGTCGTGGGACATCGCGGCCTCGTCGTGGGGA
GCAGCGGCCGGCCCCGGGGACTGCTGCGGGACTGCTCGGGCGCGGCAGCGCCCTCAAGGCGA
GTGTGGAGCTGGACCAGCTGCCTGAGCAGCCGGGTTCCCGCGCCTCCTCCGGGGAGGGGA
GTGGGACCGGCTTCCAGCCTGTGCGGCCGTGCCCGCGGCCCACCTTACGGAGTCGGTGGGCGGCGCCGGGACAGCGGC
GCCTCGAGTCGCCGAGCCCGGGCGAGCCCGGAGAAGCGGCCGTCCTGCAGCGGCCGTCGTGCGGTTGTGAGATGGGAAGCACGAATG
AAACTCTCAAAGGGTTGCACAAGGTTGCACAAGATGATCGTCCAGAGGAACGAATGATCAGGGAGAAACTGAAGGCAACCTGAAGCTGAGTCTCCAGGA
GTTGGAAAGGAGAAATAGGCGAGGGCCGTGGTGGTAAACCAATCCAGTTAAAGGAGATGGATCTGAAATCACTTAGCAGTGAGTCTTCAGGA
GAGGTCCAGCAAGTGCGGCTTCACCAGTCGGACGCAGGTCGTGGTAAACCAATCCAGTTAAAGGAGATGGATCTGAAATCACTTAGCAGTGAGTCTTCAGGA
AGTAAGGAGAAAAAGAGTTTCCCCAGGTGCCTTTCAGAAGTGGGCAGAATGCAGAGTGTACTTACTGCGAATCGCGAAGGCCTAACTCTTTCCTGATTGGAGGAGACAGCCCAGACC
GGAAACAACCCGGGTGTTAACAAAGTGATGGGCCAGACTGTACTTACTGCGAATCGCGAAGGCCTAACTCTTTCCTGATTGTGATGCTCGGGGTGTTTCAACTAGAACC
ATAAATACCGGGTGTTTATTGGGATGAGAAAACTTTAAAGATTTGAGGTTGAGAGTTTGGTGTGTTTCAAGATCAAAGCTCCATC
TTCAGACCCAATGTTATGGGAGAAAACTTTAAAGATTTGAGGTTGAGAGTTTGGTCTCCGGGGTGTTTCAAGATCAAAGCTCCATC
TCGTAACACCATCCAGAAGTTTGTTTGTTTCACGCATGTCAATCATCTAGTCATCTAGTCTTCACAATTCTCATCTAGTTCAGAAAACAGCATAAAGGATGAA

GAGGAACAGAGATGTGTCCTATTTGCTTGTTGTGTTGGCATGCTTGATGAAGAAAGTCTTACAGTGTGTGTGAAGACGGCTGCAGGAACAAGCTGCACCACCACTGCAT
GTCAATTTGGGCAGGAAGAGTGTAGAAGAAATAGAGAACCTTTAATATGTCCCCTTTGTTAGTATCAAGTGGAGATCTCAAGTGGAGATCTTCTACAGCCCACGAGTTGTC
AAGTCCTGTGGATTCCCCTTCTCCCTCAGAGCTGCACAGCAGCAAACCGTACAGCAGCAGCAGCCTTTGGCTGATCACGAAGGAATCAAGAGAGCAATTTAA
CCTTACTCATTATGGAACTCAGCAACTCCCTCCTGCTTACAAAGATTTAGCTGAGCCATGGACTGAGTGTGTTGGCAAATGGGGGAGAGCACTGGAAATTCTGGGGG
AGAAACTGGAATGTGGAAGCAGCCCGAGTGGGGGAGCCCGAGTGTCTTGCGTGCTTCCCAGAACATTGAGAGCCATGGTGGTGGAGGCATGCTGCAGCGGTTCTGTCAATGG
CAGCAGTGGAAGCAGCCCGAGTGGGGGAGCCCGAGTGTCTTGCGTGCTTAAAACATTGAGAGCCATGGTGGTGGAGGCATGCTGCAGCGGTTCTGTCAATGG
TCTGTGCTGACCCTGTCTACAAAGTGTAGCTGCTTTTAAAAACATTGAGAGCCATGGTGTATATCCTGCCACAGTTTAGCGGAAGAATCAAAC
TTCAGAGACTTCTCCAGCCAGTTGTAGCAATCCAGTGGAGACACCATTGTGCTGCTCAGTTCAAATGTGCAAGTCAGCTGTCCATATCAACACTGTTGGAACTGT
GCAAAGGCCAAGCAGGCAGGAGAGTTGGCAGTTGGCAGAGAAATACTAAAAGCTGGATCCATTGGTGGTGTTGATTATGTCTTAAATTGTATTCTTGGAA
ACCAAACTGAATCAAACAATTGGCAAGAACTTCTTGCGCGCCTTTGTTGTCCTGTCCCCTCTAACCTTTGCTTCCTGCGAATTTATCCTGAATTTATCCTCATATGTCAG
TACTGATGTTTCACAAGCTGAGCCGTTGTTGAAGTTCTGCAAGAATGTTCTGCATTACTACGCCCTCTAACCTTTGCTTCTTAGAAATGTGAGTGTTTCCAGTTCCA
GGCAAACTTCCACCAGGATGGCGTCGCGTTGATGGCTATTGCAGATGGCGTAGATCATGCCGAAGCCATCCAGTTGGCGTAGAAGACCTTGGATGGTCAAC
CTCACTTCACCAGGATGGCGTCGCGTTGATGGCTATTGCAGATGGCGTAGATCATGCCGAAGCCATCCAGTTGGCGTAGAAGACCTTGGATGGTCAAC
AGGACAGCTTCTTGCAGGCATCGTTCCCAACACTATCTGAAACCACAGAGAACAGTTCCCTGAGTGCACAATCATTAGAGAAACTGGAAAGGA
TTAATGTGCTACAAATTGAGTGCCAGTTCAGAGGACATTTCTGAGGAGACTGGCCAGCATTCAGTTCAACACACAACAACAACA
GCAACCAAAGCCAAGTGTCAAACAAAGGCAGACCCCACAGTCAGTGTTGAACCTCAGTGTTCTCCTTATCTCACTTAATGTTCCAGCCTTGT
CAACCCTTCTTCTTCTACCCATCTGTACCAGTTGTCTCTTACACATTTTCTCTACAAAATTCAGCATAGACTTCAGGGATTCATTCCCTGCAACAACTGTTACCTCTCGAGACCTT
TCCTCAAACACAGCGCAAGTTTTCTCTACAAAGCCAAAGCTTTTCTCTACAAAATTCAGCATAGACTTCAGGGATTCATTCCCTGCAACAACTGTTACCTCTCGAGACCCTT
GCCCTCCAGTAACATACACAGGCCAAAGCCATCTGAACCTCCGAAAGCCTTGGCTGTAGCAGCAATAGGGAGATCCCTCAAAAAAATAGCATGACCCCAGTGACTGAGGAGAAA
ACAGTAGTTCAAATGTGTCAATACAGAGCTCAACTCCAGTATTGGCTGTAGCAGCAATAGTAGTGCTTGAAGCATCTATGCCCCAGTGACTGAGGAGAAA
TGCAGATTAGAGTGTCAATAACAGAGCTCAACTCCAGTATTGGCTGTGAACTCTCTGAAGCATCTATGCCCCAGTAACTTTTAAGTCAGAAGTT
GCTGTCCCTGTCTCCTGAAAAAGGCTGCAATGCCATTGCAGCTGAAAAAATGTCAAAGATGATGCAAAGATGCAAAGATGAAGATCATAATCAAGATGCAAAGAAGATATCATCATTATTCAAC
AGGATACACACCAGAGACTCTACCAGGACATACCAAAGCAAAACAACCGTATAGAAGCAAAACAACACCGTATAGAAGAGGTCAACAGATAGGCCTTGGAGCATT
TTCTTCTTCTTGTTTATCAGGCTCAAGATGTGGGAACTGGAACTGAACCATCTGAATCAACAACTTTAATGGCTGTTAAACAGTTTTAAACAGTGTCAGAAAAAACACATCTTCTGAGCAAGAAGAAGTAGT
AGAAGCACTAAGAGAAGAGATAAGAAAAAACTGGGGATCGGTGGCTCATTGCTCATTTGCTTGGAGCCACGTGTGAGAGAGCAATTACAATCTCT
TCATTGAATGGATGGCAGGGGATCGGTGGCTCATTTGCTTGGAGTAAATATGGAGCCTTCAAAGATGTCAAAGATGTCAGCAGTTATTAACTACAGTTACTCCGTG
GCCTTTCGTATCTCCATGAAAACCAAAAATCATTCACAGAGATGTCAAAGGTGCCAATTGCTAAAGGTGCCAATTGCTCAGAGATTGCGAGATTTG
GAGCTGCAGCAGGTTGCCATCGCAAATGCTGCAATGCCATGGTTCAGGAGAGTTCAGGGACAATTGGGAGGAGTTACTGGGGACTACTAAGAGGT
CAACAGTATGGAAGGAGCTGGTGTGATGGTTGGCTGTTGTGGTTGTGCTATTATAGAAATGGCTTGTGTGTCAAAAAACACTCCAATCA
TCTTTGCTTTGATATTAAGAGTTTCTCAGCTCAAGATGTTCCATCGCAACTACTGCTCTTCACTTCACTTGTGTTCGTTGTTAGAACTTC
AACCTTCAGGAGGACACCTTCCATCGAAGCAGCTACTGAACCTGAAACTTATGCCAGTAACTATTGCCAGTTGAGTAGGATG
CTCAACAGAGAAAAAAACTTGTGGGGAACAATTTGATCTGTACTCAGTCTTTACTGGGCCATGATGCCAAACTGAACCAGCTATGAAGCAGGATG
AAGTATGTATTGACAAATCGATCTGTACAAATCTGTTATGCAAAAGCCCAAACTGTAACCTGAAACCTTAAATTGCTGCTTTCAAAGAAACTGGCCCTA
GGTGAACAGGAAAAAACAACAATGAGAAGTTTGACAAATCACACAAGCAGGGGGTTCTGCATTCCGGTTCTGCTATAAATCAGT
GATCTATTTTAATATTTCAATTACTTCTTCCATCGATGTCAGAGTATCATATATTTTAAAGTGATACTTTCACTGCTATAAATCAGT
ATACCAGGGCAGTTGTGGTCATCTGTTTACTGTGGGCCCATTCATTTCATTCCTTCATTTTTTTCATTTTATTTTCTTAA
TTTATTCAGGGAAAGCTGATCTTTTTTTTCAACCAGAAAAAAAAAAACCTAGTAGAACTAGAGGTTCATTAAAATATCTTGTCTATTGTGAGTTTTTAAA

>MAP3K8_H_SEQID#NA_44

CAGGATGGCCGCCATGTTTTACATGGAGCACACCGAGGGCTGATGCCTCCTTTACCAGACCACTTCTCAGAAAAATGCAGCAGCAGACTTTGTTGCGCATGTGCCT
GACCAGGGACCAGCATGAGCGACCTTCTGCTCTGAAGCACTCCTTGGAGAGAAGTCACTGAATATACATCCAGATCAAGAACTTTCTTCCCCAGTTCC
ACTGCAGATGCTCCCTTGCTTAATTGTGGGGAAATGAGGGATCTTGTTCCCCACTGAAAATTCAGTCAGTTAAGCAGATCCTATGGAG
TCATTAACTGAAAGTTGCAGTTACATATTAGCCTCCTCTCAAGTGTCAGACATTATTACTCATAGTATCAGAAAACATGTTCTTAATAACAACAAAACTATTT
CAGTGTTTACAGTTTTGATTGTCCAGGAACTACACATTCTCTATTGTTTTATATGACATTTCTTTTTTATTTTTGGCCTGTCCTGTCAATTTTAATGTTGTTAGTTTA
AAATAAATTGTAAAACCCCCTTT

>Pak5_M_SEQID#NA_45
ATGTTTGGAAAGAAGAAGAAGGCGGTGGAGATCTCAGCCCATCCAACTTCGAGCACGTGTACACACAGGCTTTGATCAGCATGAGCAGAAGTTCACAGG
GCTGCCCCGCCAGTGGCAGAGCTCGAGAGAAGAGGCCTGAATGAGGAATCAGCAGCAGACCCAAGCCCCTGATCAGCGCATCAGTCCATCAGCCTGGAGCCCCA
AAACCATCGTTCGGGGCAGCAAAGGTGCCAAGGTCCTCACTCTGCTCTGACGAATTTGAGAAACATGTCGGTGACGCGCTCCAACTCCTAGCGCA
GAGAGAGTCCACCACCACCTGCCCGTGCCCACCAGGAGAATGGAATGCTGGAGGAGCGGGGTGGGGCCAGAGAGAAATGCCCCTGACAAAGCTGGAAGCAG
AGCCCGGGCCACAGGCCACAGTGAGGCAGTGCAGTGCCACTCTCTGGGGCTGATGTCAGCACTCTCAGCCTGCTGGCAAAACTAGCAGCTG
AGGACCCCAGGAGGCCTCCCGGATAAGCGCCACTCTCTGGGCCACCACCCGCGGTGCCCAGGGGAGCCACCATACCACCATGGCCCTTCAGCCACA
GCAGGCCCTTTAACACATACCCACCGGCTGACACGGCCCTCCTCCCGGCCTCCACGGAGTTCTGGTGTCCCAGCCAGGCCCCCAGT
GGCCCCCCCAGCACGTGCCCTGCCAGCTGGTGGTGGGAACCCTGGCCCTGTATCCTAGACAACTTCATCAAGACGTGAACTGCTCTTCAATGAGGTGGTGATC
AGCAGTTCCGGGCTGCCGGAGTGGGCCTGCAGCTCAGGCAAAGTGGTGGGAGATGTGGAGAACTGGTGGTCAAACAGTTACCGGCGCGGCCCT
GCAATTGCCACTGTACGGCCAGTCACGGAGAACGTGGGTGGCTTCATGAGTTCCTGAGTTCTCTGAGTTGGTGTCAT
ATGCGGGGACTACCGGCACGAGAAACGTGGCCGCCGTGTGCTCGGTGACTTGGCTTCCTCCAGGTGTCACGCCCCAGGGTGTCAT
CACGGATATTGTCACCCACACAGAGTGACTCTATCTTGCTCAGGCCATGATGGGCCGGTTGTGCCAGGGTGAGCAAGGAGGTGCCTCG
CCACCGTGACATCAAGAGTCGCTGTGGGCCACCCGTATTGGATGGGCCCGGACTCATCTCCGGAGCTCATGGGCCAGGAGGTGATGGT
GAGGAAGTCGCTGTGGATGGGGACCTGATGGGGAGCCCCTTACTTCAACGAGGCTTCTTGGATCGGCTCTGGGGGTGATGGT
GATCGAGATGGTGATGGGGACCGTCTCTGAAAGGCTTCTTGAAGGCTTCTTGGATCGCCCTCAAATCTATGAAGATGATCCGGGACCAACCTGCCCGAGCTGCTGAAGAACCTGC
ACAAGGGCGTCACCGTTCTGAAAGGCTCTTGAAAGGCTCTTGGATCGTGCCCCTGATGCGCCCGATGCGCCAGCACCGAACCGAGAAGCACCCGTTCCTCA
CCAAGGGCGGGGCCACCAGCATGCCCAGCCAGCAGATCGTGCCCCTGATGCGCCGCAGCACCGAACCAGAGAtag

>STLK6-rs_H_SEQID#NA_46
TCGCGGCGCTTCGCCGCCTTCCCCGGCCCTCCCCGGGAGCGGCCTAGAGCACTCGCCTGCCTCGCCCTCCGCGGAGCAGGGCTCTGGCGCCGGCCCGCCCCTGTCCGCAC
CGCTGGCAGCCTGAAGAGAGTCGCTGGCCATGGTCGCTAGGTAGGATATATCTGCATCTTGAAAGGAAGATAAAACAAAAGCCTTCTTTGAATAGAT
GGATTTTGTCACTTTCTGTGTTGAACTAAAGTGAATTCAAATGTCTCTTTGGTTGAGCCAACCCTTTCCTGTCAATCCTCCATCTCCATCTTGACTTCACCAAACTGAAAAAACA
GTCTGAAACCAGTATCCATCAATATCTTGGTTGGTTGATGAGCCAAACCCTTTCCTGTCAATCCTCCATCTTGACTTGTTCCACCAACGTTTCT
CACTACGAGCTCCAAGTAGAAACTGCAATGAAAGACGGCCTGAAAGCTTTACAGAAAGAGGATTTGACAACTGGTAACTATACAACTTACTGGACA
ACAAATCTGAAAACTGAAAAACTGCAATGAAAGCCTGAAAGCTTTATGGCCGTTATTTCTCCATTTTATGGCCTCATGGTTCAACTCTTGAGGAATGAGTGAAA
GTTTTCACTGTTGCAGTTGCAGCTGCTTTGGGCTTTGGTTATTTCTGAGCGTGATCTGTGTTCAGGGTGTGTTAAGCATGCAGCAGTATTAAAGCCAGCCATATCCTCATTT
CTGGTGATGGCCTAGTGACCCTCTGGCTGAGTCCAGAACTACTGAGACAGGATTTACAGGAGTATTTACAGTGTTGGGAATTACAGCATGTGAATT
ATCAGTGCAGCCGTGGTGCAGGTGCCTTTCCAGGACATTGCATAGAACTCAGATGCATAGAAGCTCCTCTTATAGCCCATTGGATATCAGTATTTTCCC
AGCCAGTGGGCAGGTGCCTTTCCAGGACATTGCATAGAACTCAGATGCATAGAAGCTCCTCTTATAGCCCATTGGATATCAGTATTTTCCC

TCAATCAGAATCCAGAATGAAAAATTCCCGGTCAGGTGTAGACTCTGGAACTCACACAGTAAATAGTGACC
GATTACACACCACCATCCTCAAAACTTTCTCCTCCTTTAGCTTGGTACAGCTCGTTTTGCAACAGATCCTGAGAAAAGGCCATCAATACCATTGCCTCCA
ATTGTCCCATGTTTTCTCAAACAGATGAAAGCCAGGATTCAATACTTTCACGGTTGCCTCCTGCTTATAACAAGCCATCAATACTTGCCTCTATATACTTCCA
GTGTTACCTTGGACTGAGCCAGAATGTGATTTTCCTGATGAAAAGACTCATACTTGGTGCACTGGAGAATCTATTATTTAAAACCACTCT
TTTCTCCTTGCTGCTTTTTCTTCTGTATTTCTAGTTACAAAATACGAGAATTATACTTGAAAATACAGTTGGTGCACTGGAGAATCTATTATTTATTCTGAATT
GTTCAAAGGGGCACCAGTTTGTGTAGTCCCTCTGTTTTGCACAGAGTACTATGACAAGGAAACATCAGAATTACCAATCTAGCTAGTGTCATTTATTCTGAATT
TTTTTCTAAGCTGTGACTAACTCTTTTTATCTCAATATAATTTTTGAGCCAGTTAATTTTTCAGTATTTTGCTGTTCCTTGGGAATGAGCCCTCAGAGGAC
AGTGCTTCCAAGTACATCTTCTCTCCCAGATTCTCTGCCTTTTTAATGAGCTATTGTTAAACCAACAGGCTAGTTTATCTTACATCAGACCCTTTCTGGTA
GAGGGAAAATGTTTGTGCTTTCTTCTTTTTCTTTTTCTGTTAATAATCTATGGTAACACTCAACTTAAATTCATTGCACCCGTGGGAGCTCAATATACAGA
AATTGTAAATTTGCTTTTTTTTTAAAAAGGGGGCTAAAGTAACACTTTCCTACTTTGTAAATTATAGATCTTAAATTCATGCACCCGTGGGAGCTCAATAAAG
ATATACTGAATTGA

>MAP2K2_H_SEQID#NA_47
CTCTTGGACTCCGGCTGCAGCGTCAGCCTCCCTCGGCCTCAGCCTCGGCTCAGCCCCAGCGCCCTCGGCTACCCTCGGACCAGGC
CCGCAGCGCCGCCCGCCTCGGCGCCCCGACGCCGCCTGGCCCCCGGCTCGCGTAGGCGGCCGAAGCGCTCCCGGCCCCACTACG
GGCCCAGGCTAGAGGCGGCCGCCGCCACGGCGGGGCGCTCCGAGCGGCCGATGCTGCCGGCGGCTCACCCTACCATCGCC
GAGGGCCCGTCCCAACCAGCGAGGGGCGCTCCGAGGCAAACCTGGTGGACACTGATGCTGGACGAGGAGCGGCAGGAAGCGGC
TGGAAGCCTTTCTCACCCAGAAAAGCCAAGGTCGGCGAACTCAAGGTCGGCGAACTGACTTCGAAAGACGATGACTTCAGAGGTCAC
CAAAGTCCAGCACAGACCCTCGGCGCAACTCACGTACTACATCGTGGCTTCACGTGGCATTCATGACCAGGAGAATCAAGCGGCCATCAAGGAG
TCCTGCACGAGTGCAACTCACGTACTCACGTGGCTTCTACTGTGACAGGAGCCTTCTACTGTGACAGGGAGATCAGCATCTGAGCAGCACATGATGGCCGGCTCCC
TGGACCAGGGCTGAAACGTGCACGGGGCTGAAACGTGCAAGAGAGGCCAAGAGAGCCCTCAAACATCTCGTGAACTCTAGAGGGAGATCAAGGCAGCATTCGGGTTGGCGTACCTCCGAGAGAAGCA
CCAGATCATGCCAACTCCTTCGGTGGGGCCAATGGCTGGCACGCGTCCCACTACATCCCCGGGCGCCCCAAGCGGTTGCAGGGCACACTTACTCGGTCGGTCAGTCGGTCATCTGGAGCATGGACCTGT
CCCTGGGTGGGAGCTGCATCGAAAGTGGGCCATGCGTGAACTCCTGCTCTAAGCTCTAAGCGCACACGGTACCCCGAGCGTTGGCCCCAGGGAAGAAGGA
GAGCCTCACAGCAGCAGCATCTCTCCTTGGGCAGCCTTCACAAACCACGCCTCTCATCTGAAAGAGCCGGTTGCATAAATGCCTCATCAAAAACCAAGGGCTGAACCA
TATATTGTGAAAGAGATGCTCACAAAACCAGCCCACCGCCTTGTAACAGTGCGGTTGCCTCAAAGCGGTTGCAATGCCTCAGGCTGGGGCTGAACCA
GCGGACCTAAAGATGTCTCACAAAACCAGCCGCGTTTGCCATGTCCTCTAGTTGTGTAAAAGCGGATTTTGCCTGTAGTTGTGTAAAACCCTGGGGCTGAACCA
GCCCGGCACACCCACGCGCGCCTCCAAAACCACCCCCACCCCTCCTCACCCCTGGGGCTGGGGCTGGACCA
GAGGACAGGCTGGCGCTCCTCCAGGGCGACCCGACCTGTGTGTGGTCTCAGAGCCGACCTGTGTGTGGTCTCAGAGCCTCTGGTCTCTTCCTAGGGTTCTCAGAGGAGAAA
TGCTGGTCCCTGCCGGCGTCCCTCTAGGGACGGGGCGACCGCTCGTGTGTGGTCTCTTCCTAGGTTACAAAACAAGGGGAGAGAAA
AACC

>CCK4_H_SEQID#NA_48
CGGGGACTCGGAGGTACTGGGGCGGCGGCCTCGGGACGGCCTCGGGACTCGGGCGCGGCCTCGGCGCGGCCCGCGC
TCCGGTGCGCTCCGTGCGCCGGAGCGCAGTCTGCGGCCCTCAGCTCCTTTCCTGAGCCCGGGAGCTCGCGC
GGGGAATCCCGGCCAGACCCGCCGGTTGCCTCGTCCGGCGCGGCCGCAGCGCCATTGTCTTCATCAAGCAGCCGT
CCTCCCAGGATGCCACTGCAGGGGCGCGGGCGTTTGCCCAGGCAGCAGCTGGGCCTGCAGGGTCGCGGCCCTGTCC
AGGACACGGAGCGGCCGGTTTCGCCCCAGCGTTGCAGGGCAGCTGGTAACATGTGTACTGGGCCGGATGGGCCCGTGTCC
GTCACTGGAGGAGAAGAGCCCGCAGTGCCAAGCGCCTCCTTCAACATCAAAATGGATTGAGGCAGCATCCAGCCTCGAAGCATCCAGCCTCGGAAGCTGAGAT

>LMR1_H_SEQID#NA_49

```
GGGTGATGCGGCGGCTCGGCGGCCGGCGGCGTCCTGAGCGCAGCGCGCCTCCCGGGATGCCCGGCCGCTGGCCCGGCCATGCTGTCGTCGTCCTTCTTC
AACCCCAGCTTGCGCTTCAGTCGCCTTGCACCCGACCCGGCGCGCCCGCTTCAGCGAGCTGTCCTGGCCATCCTCCCTGCGTGTCGCCTGTGTCTTTCTCCG
GGCTCTTGCGCGTCATCGTCCTCATGCTGGCCTGTCGTTAAGAAGGCGGTATCGGGTTCAAGGAGTTTGAGAATGCGGAGGGGACGAGTACGCA
GCCGACCTGGCCAGGCTCAGTGCGCACGGCACAGAACGGGCCGACGTGTACGTCTCTGAAGGAAATCGGCCGTGGCTCCTTGCCCATGGCCAAGCAGCC
TGGGCCGGCTCAGTCAGTGCAGCCTCCTCAAGTCCACAGAGTGGGCCGGCACAGCCTAGTGCCAGGAGCTGCAGGAGCAGATGCAGTTCGTGGAGGAGGTGCA
GGGAGGTGAACTCTGGCATCGCCAGGTGCTGGTGGCCAGTGCCTCCAGGGTGGTGGAGCTAGTGCCAGGTGACGCCTAGTGGAGTTCTGCCCACTGGGGG
GCCCTACAGGGCTACCTGCCGGAGTCCCGGCGGAGTCCATGGCTCCGCGGAACTGCCTGCTGACCCTGCTCACGGCTCGACCTGCTGAGGTGGCCTGTGTGGGGTCCTG
ACCTCAAGGGGCTACCTGGAGTTCGTGCACAGCGGACCTGGCCCTGCCGAACTGGCCTCTGCGCTGTGGGTGCCTTCTGCGCCTGCTGACGGTGAAGATTGGTGACTATGGCCTGGCTCAC
CACCTTCATGCAACAATTTCGTGCACAGCGACTACTTCGTGACCTGTGGGTGCCTCTGCGCTGGACTGCGCCAGAGCTGGTGGTGGACGAGGTGCATAGCAACCTGCT
TGCAAGTACAGAGAGACCAAGAGCGGGAAGTGTTGGTCCGGAGCAGCTCAGCTGCCCAAGCTGCCTGTTTGACCCATCTGGGAGCTCTTTGACCTGTCCAGCCTGGGACC
CGTCGTGGACCAGACCAAGAGCGGGTACACGGTCCGGGAGCAGGGCCCACAGCGTGCCTGTCGGGAGCTGGTACGGATTGTCAGTCGATGCAGTTC
AGCAGGGTGCTGGCTGCGACCGCGGCAGCGGGGCCCACAGGAGGTGCCACCCTGTGTGCCTACCGTGTCTGTGCCACGGCGCCACCGGCGTGGTACGGAGGAGTTTGA
TGCTGGCTGCGCACGGCCGAGCGGGGCCCACAGGAGGTGCCACCCTGTGTGCCTACCGTGTCTGTGCCGTGGGCGGTGTGGGCGTCGCCCGTGCCT
ACGGCGCTGGCGCTCTTGCGCTGGTGGAGCAGTTCGGCGGCGAGGCCTTCCCGGCGCACGGCGACGTGCTGACGGGAGACCAGCCAGGCCTCAATTTGAG
CGTCCTTCCGCTGCTGGAGCAGTTCGGCGGCGAGGCCTTCCCGGCGCGCGACGGTGAGCCCTGGCCACGCTGCACCTGCAGAGGGGCGCCACGACCCTG
TACAAGTGGGGAGCGGCGGCGCGGGAGGCCGCGCACGCCGTGGTTCCGGTGCTCAGCGCGCACCTGCTGGCGCAGAGGCGAGTACTTCATCCGGCCTAGAGAGGGGCCACCGTGGTGCCCCGTTGG
CCCGGGCGGTGGTTCCGGTGCTCAGCGCGCACCTGCTGGCGCAGAGGCGAGTACTTCATCCGGCAGCGGATGCTGGGGCGGTGTGGGGCGGTGGGC
ACTGCGCGCCCGTGCACGTCCCCGTGGGGCGGCCACCTACCCTCGCAGAAGCTTGGCGCGGCCTTCGTGCTGGGGGCGGTTAGAGGAGCGGTCTGGGGC
CACGGGCCACCCCGTGCACGTCCCCGTGGGGCGGCCACCTACCCTCGCAGAAGCTTGGCGCGGCCTTCGTGCTGGGGGCGGTTAGAGGAGCGGTCCTCG
GCGGGGCCCCTGAGTGCTGGCGGGAGGGAGGCCCCGCTGCGGCCACCGGGAGGGGAGGAGGATGCGCGCTGACTGGCGCGAGGAGGGGGGGCCCCGCTCC
TTTGGGGAGCTCAGGGGCGCGGGGGCGCCCCGCTGCCCGGGGTGTGCCGCCACGGGCCCTTCTGCCTTCGTCTTCTTGTCTGCTTTGGGCGTGGGGCCACTGGCGCGTCCC
AACGTGTCAGCCAACAACAACAGGGCGGGTACCGTGGCCTCTGGTTACACCCTCCTGGTTACACCCTCTGGATTGCGCGCTGAGGGGCTGCCCAGGGCTGCCCAGTCCAAAGCAGAC
CCCACGGGCCTCCCCAGGGCCTGGCCACTGCCCGCCACCTGCTGTTACACCCTCCTGGTTACACCCTCCTGCGCTTCCTTCCGTCCCCGGCTGCCCAGGGCTGCCCAGTCCAAAGCAGAC
ATGCTCTGCCCAGCGCCCGATGGCCGGGGCCACCTGGCCACTGCCCACTGGTGCGCGCCACCCCGTGCACGTGTGGGGGTGACCAGTAGTGGGGTGACCAGCAGGAGCCCAAGC
TTGCCACGGAGGCGCTGAGGGCCACTACCGGCACCTGGCCACTGCTCTCTCCCGGCCTCCGTGCCTTCCACTCCCTGAATGGCCGCTGTCAGCAGCAGCT
CCCCCGACGGCCCCCGATGCCCGATGGCCGCTGCTCTCTCCCGGCCTCCGCGTGTCTGTGCCGCTACTGGTGCGAGGGCCTGAGGGCCAGCAGCAGGAG
CTCCGGAGGTGGAGGCACCCAGCAGTGGAGGATGAGGAAGCCAGGCAGGTGCGGCCAGCAGCGCCCGAGGGCCTGGGACGCCGCCTGAGCGCCTGCAGGAG
GCCGGATGGTGGTTCAGCCGGTCAGCCTTCGCGGTCTGCGCCCTGGCCACTGCGCGTGCGGCCTATGGCTATGACACCGAGAACTATAGTCCCCTGAGTTTGTGCT
TGAGGTCTTCAGCCGTGGCCCTTCGCGGTCTGCGCCCTGGCCACTGCGGGGCCTATGGCTATGACACCGAGAACTATAGTCCCCTGAGTTTGTGCT
CAAGGAGGGCAGGAAAGGGTGTGAGAAGGCCTTTGCCGAAGGGCTGGAGGGCTGAGCCTCAGAGGGTGAGGCCGGAGCTGAGCCAGAGGCTCCCTCAGTG
GCCTCAAACGAGACAGAAGAATCCCTACCGAGACTCTGCCTACTTCTCAGCCGAGACTCTGCCTACTTCTGAGCCTCCAGAGAAGGCGGGGGGAC
CGAGCCCCGGCGGGCCAGAGCTGGCCGTGCTCCCGCTGCTGGCCGCCACTGCCCCACTGGTGCGCGCCAGGTGTCCGGGTTTCCGGGGAGGCACAAGGCTCTGGCCC
CGGGGAGGTGCTGCCCTCCCCACTGCTTCAGCTTGAAGGGGTCTCCCAGTTTTCCTGCTGAACCCCCAGATCAGAAGGCAACAGCTCTGAGTTCCAGGGGC
CAGCCAAGGTGCGGCGACCTGGAAGCGCAAGAAGGCCGTGTCCTGTCCCACCCGATGGGGGCCCACAGGGACAGTCGGCTGTACAGGCCTACAGGCTCCGCCC
CCCAGGACTGTTGTCAGGGCCGGCCGCCACAAAAGCGGATGGGGGCCCACAGGGACAGTCGGCTGTACAGGGAAAGCCCCACCCGGAGCTCGGG
TTGGAGGGCCGGCCCGGCCGGGGAGGAGGAGCGGCGTGGGGTGCCGTGAGAGCTGAGAAGATGCCTCAAGATGCTCAAGATGCTCTGCCAGCTGTCCGAGA
AGCGAAGAGAGGAGGCCGCCGGCGGAAGAATCGCCCCGCCAAGGACACGTCTCCCCACCGTTCCTTAGGGGGAGGCCGGGAGCAGGCCTGGCTCGGG
CCTTCTGCGAGGACCTGGAACGCAAGAGAGGCCGGGTGTCCCCTACGTTCCTTAGGGGGAGGGCCCAACCGGCCCAGCGTTCCCGGGAGCTCGGG
AGCCCCTTCCCGGGCGGCCAAGGAATCGCCCCGCCAAGGAGGGCGGGAGCCGGCAGCGGCCCCAACCGGCCCAGCGTCCAGCAGGCTGATGGCTCCCCCA
```

AATGGCTCCACAGCGGAAGAGAGGGTGGTGGCGTGGGACGAGCGACTTCCGCTGATGACGGCCAAGGCAGCCTTGCGCCATGGCCCTAGACCCGGCCGC
ACCGCGCCCGGCTGCGCCCACGCCGGCCTCCCTTCTCGCGCTTCACGGTGCCGCCGCGTCCCGGCTTCTCCATCACGCACGTGTCTGACTCG
GACGCCGAGTCCAAGAGAGGACCCTGAAGCTTGGTGCCGGGGGTGAGAGTAAAGAGAGGCTTGAGACCTGGCAGCTCCTGCCCCTCAAGGCTGGCGTCACCGG
AGCCCCTGCCAGGCAGCCAGCAGCGAGGATGGTGACCGAGAAGGTGGGACCCCATGCTGTTGGCAGCAGAATTCGTCGTTGGCTTGCTGCTAGCCCCT
CCTGGAGAAGCCCTGGAGCCCGTGGGATGGTGTAGAGAGGCCCTGGTGTCTGTACACACATACAAAAGGGGCACGGCGCCCCCTGGCACCCTCTGCCCTGCCCTGCC
GGGGGCGGCCTGGAGCCGCCACAGTGGGGTGTCTGTACACATCCGGTGTGCTCCTCCGGTGTGCTCAGTCGCTGCCCACCCTTGCTGCTGGTGTTCCTGGTTGCCTGTGTGCTCCTGTGGTCC
TGGCCTCGGAGGAGGACCCGCATGCCCATGCCAGTGCTTAAACCAGGACACTTAAACCAGGACGAGGCATGCCCGAGACACTGGCAGGTTT
GTGAGCCTCTTCCCACCCCCTGTGCCCCCACCCTTGCCTGTGTTCCCTTCCCTTCACTGACTAGACCCGTAGCCCCACCCTTCCCAGGGAACAGGCTGCTCCCACCTGGTCCCGCTGT
CTTATCTCAAAGTCCGTGGCTGTTTCCCTTCACTGATCAGGGGTGGAGGGCCTTCCAGGCTTCCCCAGCTGTGCCCAGTATCTGTGTCCCGCTGCCTGC
GGCCACGGTGGGCAGCCAAAAGATCAGGGGTGGAGGGCCTTCCAGGCTTCCCCAGCTGTGCCCAGTATCTGTGTCCCCGCTGCGAGGACCGTGC
AGGCAGCTCCCCTCTGTGGGGCAGTATCTGTGTCCCGCTGCCCAAAGGAGAGTGGGGCCCATGCCCGCAGTCAGTGTTGGGGGGCTCTGCCTAC
AGGGAGAGGGATGGTGGGAAGGGGGTGGAGCTGGGGGTGTCCAAAGAGATGTAAAGTTTATTGTTGCCAAAGAGATGTAAAGTTTATTGTTGGTTGGAGCGAATGGAAGTT
GGGTGATTTTAAGTTATTGTTGCCAAAGAGATGTAAAGTTTATTGTTGCTTCGCAGGGGATTTGTTTGTTGTTTGTTTGTTTGAGGCTTAGAACGCTGGTGCAAT
GTTTTCTTGTTTCCTTGTTTTTTAAGAGAAATGAAGCTAAGAAAAAG

>RYK_H_SEQID#NA_50
CGGCTCGGGGGCTGTGAGCGGCTCGGGGGCCGGGGGTGGGCGGCGGGGTGCGGCGGGCGGGGGGGGGCCATGCGTGGGG
CGGCGCGGGCTGGGGGGCCTGGGGGGGCGGGAGTTGCCTCCCCGGGGGCCGGGGGCCTGCGTCGCTGTTGCCGGCGC
TGCTGCCCCGGGGCGGCGGCTGGCTGGGCGGTCTGTGTGGTCTTCTGTGATGCGAATGCCCTAGTGCCTCAGTGAGGCACG
AGGTGCGCGGCCGGCTGATCGGTCTTGATGGCAGGAAAACTTTATTATGTGAGAAAATGAATAAAGCTTCCAAGTGGATATGCCCCCCAGTGTCAACAT
AAATTTCCTGCACTTCACCTGGCCATGGCAGTTCCACGCGCACTTTATCAGTGTTTCGGGTTAGAGCTTTATCAGTGTTTCCTGTACTGGCAAAGTTAGAGCTTTATCAGCTCAAC
TTCTGTTCAGGGGGAAGTTCACGCACTTTATCAGTGTTTCCTGTACTGGCAAAGTTAGAGCTTTATCAGCTTTATCAGTGTTTGAAGTTATGATACTAAATGCAGCTCAAC
TTGACAGTAAATTCTTCAAAAATTCTTCAAAAATTCTTACCGTCTTAAAATTTAACGAAGGAAAAATGTCTACAAAAAACTGAAGAAGAAGTAAAAACTTCAGCCTTGGACAAA
AACACTAGCAGAACTATTTATGATCGTCGTACATGCAGTAGTGACCAGCATTAGTCCAGCAGTAGTCCAAGGGCTGTCTCAGCCAT
TAGCAATAATATTAGCTGTTTTGCACCTTCATAGTATGAAAAGGATTGAAAAGGATTGAACTGACAGCATTAGTCCAGCAGTAGTCCAAGGGCTGTCTCAGCCAT
CCACCCAGACGACTCAGTATCTGAGAGCAGAGCAGACAGCAGCAATGCAATCCTATCCTGAGAGAGAGGATAACTCTAAAAGATGTACTCCAAGAGACTTGAGAAGT
GTCACTCTTTTTGGAGGCAAGGGCAAGGTGAAGGTCAAGGCAAGATATAGCAATATCCAGAGAGGATATAGCAATATCCAGAGAGGATAACTCTAAAAGATGTACTCCAAGGTACTTTTGGGCGTAT
TTCCATGGGATTTTAATAGATGAAAAAGATCCAAATAAAGAAAAAAACATTGTCAAAAACAGTTAAAGATCCAAATTCTGAAATTCAGGTGACAATGA
TGCTCACTGAAAGTTGTAAGCTGCGAGGTCTTCATCACCGAGTCTTCATCACCGAGTCTTCATCACCGAGTAGTCATGTGTGTAGAAGAGAAAAGCCCATGGTGATATTGC
CTTACATGAATTGGGGGAATCTTAAATTGTTTTTACGACAGTTAGTCCATGTGTGTAGAAGAGAAAAGCCCATGGTGATATTGCCAAGACCTGGTACACATGG
CTATTCAGATTGCCTGTGGAAATGGACAATGCCCTCTCCAGAGACTTGTTCCCCAGGAAATGGACAATCACTGTCATTGATGACGAACTGTGTTGGATGGCTCTTGAAAGTA
AGATCACACAGAACAATGCCCTCTCCAGAGACTTGTTCCCCAGGAAATGGACAATCACTGTCATTGATGACGAACTGTGTTGGATGGCTCTTGAAAGTC
TGGTTAATAAACGAGTTCTCTAGCGCTAGAAAGATGGTTAACCGAATAGCCCAGCCAATCAACTGTCCTGATGAATTATTTGCTGTGTATGGCGTGTTCTGGGCCTTAG
CCTTTGAGATGGCCGCATACCTGGAAAGATGGTTAACCGAATAGCCCAGCCAATCAACTGTCCTGATGAATTATTTGCTGTGTATGGCGTGTTCTGGGCCTTAG
ATCCAGAGGAGGAAGGTGCCTGTGTCGGGGCTCACTGAACTTTCATGCAGCCCTAACAACGCCAACAGAAGCACATTGTCTTCTCCAGAACACCGTGC
CTTAGAAATGCTTTAGAAATCTGAACTTTTAAAACATACTACTTTGGACTAGGGTACATTCTTACAAATAAAATAAACAGTTTTAAAATTGTTTAGACACAGATATTTG
ATTTTTTGGCCAAAATTTTTAAAACATACTACTTTGGACTAGGGTACATTCTTACAAATAAAATAAACAGTTTTAAAATTGTTTAGTGTTATTTTTATCACTACT
GAATTAGCTATCTTAGTGCCAACTGCTTTTATTTTTTACTTCATCAAGGTGATGTAAGTGACTCACCTTTAAGTTTTTTAGTGTTATTTTTATCACTACT

CTGGGAAATGTGTTGTCTTCAAGATGCAATACTTTCTTAGTAAGGAAAAAACAGCATAAAAAGATACCTGGTCTGCCTTGTACAAGAAAAGGCAATATTAG
AGGAAGAAAATTTAAAGAAAAGCTAGAGGAAAAGTAAATATTTTTAAAAAATAACTTATTAGAAGCAAACTGCCCTTGCATGGAAAACTGTTTATTTTTTCAG
TGAAAAGGAATTCTGCTTTCGTGTTTTCGTGTTTTGGGAAGCAGGAACTGAGTTCATTACTCTCTGGAAAAGGAGATAATACAATTCCAGAAAGTGAACTCATATTTCTCAAGGTTAAGA
GGGCAGATCTGTAGTAAACAATGGTGATTTATTTATTTACTCTGGAAATGTCATGCACAGAGCGGGTGTGTTGCTTGAGTGTGTCGTTTTTGTTTTGTTTTTTGTTTTTGGTTGTAGGTTAAACTGGTAAACTGTAAAATTTTGTGCTT
ATCTTCAAGGCTGGCTTAAGTATAAAAATTGTTTTTAAACACTTGAAAAATTAAAGATTTGTTTTTTATTATTATGAACAGTATTGAAATTATTTTCATAATGAAT
GATTGGTTATTGTTCTGGTAAGTCTTTGGTAAGTCATTCAACAGCCAGACATTTGTTTTATTTTCAGTCAAGTTCCAAGCCCTAACACAGTTAA
GCTGGCTCAGACTCCAGGTTCTAGTAAAAAGTTGGAATTAAATGTTATAAGGAAGTATTAAAACACTGAAAACATTTCTCCAGCAGCAGCAAGTAAGGGATAT
GTATGTATTTATGCTCAGTTTTAGTTGGCCTAAAGCAGAGTTGAAATGGCTTTCTAAATAGCTAGGCCTTCTAAATAGCTACCTCCATCTTCAGAGGT
ATATAAGGGAGAAATGTTGTGAGCAGTTGGAGGCTTTGCTGTTTTATTATTATTAACAGTTCTTTGTGTTGTTTTTTGTGTTTTTAAAAAATAGCCATGGGTAGTGATTTGTTATTTTGTTATTTTCAGTACCTGGTT
TTCCACAAGTAAGCCAAGGGCATGATTTTCTGTTACATTATTATATTTCATAACCTGCAATTCAACCTGCAACTCAACACTTGTGTATTTTCCACA
AAAAGGTCCCTTTGAACCATTTGTAACCATTTGTAATTATTTTCATTCAATCAGGGCATTCTTATTTCAGTTCAAGTGAGTGCATTGTTTAGAAA
TGGAGTTGGTTCAACTCAAGCGTCCCTTCCACTTCCTTCCAGTATTCAGGGCATTCTTATTTCAGTTCAAGTGAGTGCATTGTTTAGAAA

>LRRK2_H_SEQID#NA_51
GAGATCTGGAGATAGGCGGCTTTCATTTTTCCTGCTCCCAGTTCCCGTGGGGCTGGGGCTGGGGCTGGGGCGGGGGTGGCCGGC
CGGGGCTGCCGACTGCCGCTCCGGGGCTCCGGGGCTCGTGTGGACTCAGCGAGTCCGCTGACTCAGTTTCTTCCCGCGACTCCCGCGCCCGCGCC
GCTGCGGTGGAATCTGGTCCCAGGAGCGGGCCTCGGCGGGTTCGGCCTCCGGGGCTCGGGGCTGCACGGTCACGGTGGGCCAGCCAGCCCGGCTCC
AGCAGCCCCACGGCCCCAGAGTTCTGCGCGCGGGCCCCTCGGGGCCCTGCTGCGGGGCGGCGGGTTTTGGGGCCTGAGTGGGGGGAGGA
GGAAGCCGAGCAGGAGGGCTCCGGAGAGGGCTCCGGAGAGGGCAAGCGGGGAGCTGGGGGGGTCGCCCGGAGCGTGTGGGGCCGATGGGGCCCGC
GGGGAGCGCTGCGGGGGTCGGGGCGTGAGCTGAGCTCAGGGGGTGCCCGGGGTCAGGGGTGCGAGCAGCGTTCATGCTGGGCTGGGAGGG
CGGCGGGTTGGAAGCAGGTGCCACCATGGCTAGTGCCACCATGGCTAGTGCCAATGTCCAAATCGGAGTCCAAATCTGGAGGTCCAGGTGGTGCGAGGGTGCGAGCAGGGAAACTCTGAAGAAGTTGATAGTCAGGCTGAACAAT
GTCCAGGAAGGAAAAACAGATAGAAAACGCTGGTCCAATCCTGGACTCCTATATGGACCCAGAGTGGGAGGTGTTCAGCAGGTGCGCGAGTCGCGAGGTGGTTCACCAGTTCCTTCGTGTCACCAGTTCCTTCGTGTGTTCACCAATTGAATTCTTAAAATGTCTAACAGTTCATAAATGCC
GGTACAATGCAAAGCTTAATGGACCCCAGATGTTGGAAATGATGGGAAATGATGGAAATGATGGGAAATGATGGGAGTCAGTGTAATATGCTGGAGGAATATTTCATGT
TAATTTTTGATGCCATGCCCATCATTTCCAGCCAATGATATATTCCAGCCAATGATATGTATTTGTCGTTAGTGTCGTTTGAGAGAGTCTCAGAGGAGCAAC
TGACTGAATTGTTGTGAGAACAAGATTATATGTGGAAGTCCTCATAGGCTTACAGTGAGTTGGAAGCTATGAAAGCATTCCTATGAGTGA
TTCCCTAGCGATTCCTTGCAATAATGTGGAAGTCGTTTGCTGTTTTGGCCCTTACAGTGTTTGGCTGGCAATTAGGTAATTTTTCAATATCCGTTAAGATGGAAGTGTTTCATTGTCTGCATTGTCTGTCATTTACA
AAGAATTCAAGAAGTGAGTTGCGTTTGCTGTTTGCAATGTGGAAGTCCATGAGTTTGTGGTGAAAGCT
GTGCAGCAGTACCCAGAGAATGCAGTGGAAGATGAGGGGAAGAGATAAATCTCTTAAATCAGATTTAGAGGAAAAAG
AATGAGAATCAGAAGACCGCATGCTGGGCACTAAATATCTCTTATGGCACTAAAATCTCCTATGGCACTAAAAACAGTTACCAAAAGGAAGTTTTCCTAAAGGAAGTCCAGGAAATGGCCATTTCCCAGCTCATAGG
GAAGTGATGCTCTCCATGCTGATTGCATTCTCCATGCTGATTGCATCTTTAGAAAGGAAGTTTCTTAGAAAGTGTTAATTTCAGAA
AAATACTGTTATCAAAAGGAATAACACTTCCCTGACATTTTACATTTTACAAAAACTGGTCAGCTTTGATATAAGTGTGGCTGGCAGCTTTTATAGTGTCCTAGCAGCGTTTTATAGTGTCCTAGCAGCTTTGAAATGCTAAATC
TGGAGGGCTTCGAGCTATTTTACGTTGGGTTTTTTAATGGCTGGCATGCCTGGCAGCTTTGAACAGGTTCATTGGAAATCCTGGAAATCCTGGGAATTCCCAAAGAATCCAGAAAGAAACATATCATTACCAGTGCAGC
AGTGTTTCAAGAATGATGATATTTCACAAAACTGTTTACAAAACTGTTTATAGTGGGCTATGATGCATTAGAGAGATGTTATCCGAGATTCTCGAGAGATGTTATCCAGATTATCCCAGATGTTATCCCCTGAAGGTGGCTATGGGTGCTAGCAGTG
TTGTACATTTTCCTGATGCATTAGAGAGATGTTATCCGAGATTCTCGAGAGATGTTATCCAGATTATCCCAGATGTTATCCCCTGAAGGTGGCTATGGGTGCTAGCAGTG

TCTGGGTTTAAGTCTTATAGGATACACTTGATTACAAAGAAGAAGAATGTGTTCATAGGAACTGGACACATCTGCTGGCAAAAAATTCTGGTTCGGTTCGTGCTGGTGCATCATTCATTTGACT
AAGGATGTTGCTGAAATACAGACTAAAAGGATTTCAGCAATCTTAGCAATCCTCAAATTGTCAAGCATCTTTCTAAGCTGCTCTTTTCTAAGCTGGTGCATCATTCATTTGACT

>pMLK4_H_SEQID#NA_52

ACGAGGCTCGCGGCGAGGACGAGCTGAGCCTGCGGCGCGGCCAGCTGGTGGAGGTGCTGTCGCAGGACGCCGCCGTGTCGGGCGACGAGGGCTGGTGGGC
AGGCCAGGTGCAGCGGCGCCTCGGCATCTTCCCCGCCAACTACGTGGCTCCCTGCCGCCCGGCCGCCAGCCCCGCGCCGCCGCCCTCGCGGCCCAGCTCCCC
GGTACACGTCGCCTTCGAGCGGCTGGAGCTGAAGGAGCTCATCGGCGCTGGGGGCTTCGGGCAGGTGTACCGCGCCACCTGGCAGGGCCAGGAGGTGGCCG
TGAAGGCGGCGCGCCAGGACCCGGAGCAGGACGCGGCGGCGGCTGCGGAGAGCGTGCGGCGCGAGGCTCGGCTCTTCGCCATGCTGCGGCACCCCAACAT
CATCGAGCTGCGCGGCGTGTGCCTGCAGCAGCCGCACCTCTGCCTGGTGCTGGAGTTCGCCCGCGGCGGAGCGCTCAACCGAGCGCTGGCCGCTGCCAACG
CCGCCCCGGACCCGCGCGCGCCCGGCCCCCGCCGCGCGCCGCATCCCTCCGCACGTGCTGGTCAACTGGGCCGTGCAGATAGCGCGGGGCATGCTCTAC
CTGCATGAGGAGGCCTTCGTGCCCATCCTGCACCGGGACCTCAAGTCCAGCAACATTTTGCTACTTGAGAAGATAGAACATGATGACATCTGCAATAAAACT
TTGAAGATTACAGATTTTGGGTTGGCGAGGGAATGGCACAGGACCACCAAAATGAGCACAGCAGGCACCTATGCCTGGATGGCCCCCGAAGTGATCAAGTC
TTCCTTGTTTTCTAAGGGAAGCGACATCTGGAGCTATGGAGTGCTGCTGTGGGAACTGCTCACCGGAGAAGTCCCCTATCGGGGCATTGATGGCCTCGCCGT
GGCTTATGGGGTAGCAGTCAATAAACTCACTTTGCCCATTCCATCCACCTGCCCTGAGCCGTTTGCCAAGCTCATGAAAGAATGCTGGCAACAAGACCCTCA
TATTCGTCCATCGTTTGCCTTAATTCTCGAACAGTTGACTGCTATTGAAGGGGCAGTGATGACTGAGATGCCTCAAGAATCTTTTCATTCCATGCAAGATGAC
TGGAAACTAGAAATTCAACAAATGTTTGATGAGTTGAGAACAAAGGAAAAGGAGCTGCGATCCCGGGAAGAGGAGCTGACTCGGGCGGCTCTGCAGCAGA
AGTCTCAGGAGGAGCTGCTAAAGCGGCGTGAGCAGCAGCTGGCAGAGCGCGAGATCGACGTGCTGGAGCGGGAACTTAACATTCTGATATTCCAGCTAAAC
CAGGAGAAGCCCAAGGTAAAGAAGAGGAAGGGCAAGTTTAAGAGAAGTCGTTTAAAGCTCAAAGATGGACATCGAATCAGTTTACCTTCAGATTTCCAGC
ACAAGATAACCGTGCAGGCCTCTCCCAACTTGGACAAACGGCGGAGCCTGAACAGCAGCAGTTCCAGTCCCCCGAGCAGCCCCACAATGATGCCCCGACTC
CGAGCCATACAGTTGACTTCAGATGAAAGCAATAAAACTTGGGGAAGGAACACAGTCTTTCGACAAGAAGAATTTGAGGATGTAAAAAGGAATTTTAAGA
AAAAAGGTTGTACCTGGGGACCAAATTCCATTCAAATGAAAGATAGAACAGATTGCAAAGAAAGGATAAGACCTCTCTCCGATGGCAACAGTCCTTGGTCA
ACTATCTTAATAAAAAATCAGAAAACCATGCCCTTGGCTTCATTGTTTGTGGACCAGCCAGGGTCCTGTGAAGAGCCAAAACTTTCCCCTGATGGATTAGAA
CACAGAAAACCAAAACAAATAAAATTGCCTAGTCAGGCCTACATTGATCTACCTCTTGGGAAAGATGCTCAGAGAGAGAATCCTGCAGAAGCTGAAAGCTG
GGAGGAGGCAGCCTCTGCGAATGCTGCCACAGTCTCCATTGAGATGACTCCTACGAATAGTCTGAGTAGATCCCCCCAGAGAAAGAAAACGGAGTCAGCTC
TGTATGGGTGCACCGTCCTTCTGGCATCGGTGGCTCTGGGACTGGACCTCAGAGAGCTTCATAAAGCACAGGCTGCTGAAGAACCGTTGCCCAAGGAAGAG
AAGAAGAAACGAGAGGGAATCTTCCAGCGGGCTTCCAAGTCCCGCAGAAGCGCCAGTCCTCCCACAAGCCTGCCATCCACCTGTGGGGAGGCCAGCAGCCC
ACCCTCCCTGCCACTGTCAAGTGCCCTGGGCATCCTCTCCACACCTTCTTTCTCCACAAAGTGCCTGCTGCAGATGGACAGTGAAGATCCACTGGTGGACAGT
GCACCTGTCACTTGTGACTCTGAGATGCTCACTCCGGATTTTTGTCCCACTGCCCCAGGAAGTGGTCGTGAGCCAGCCCTCATGCCAAGACTTGACACTGATT
GTAGTGTATCAAGAAACTTGCCGTCTTCCTTCCTACAGCAGACATGTGGGAATGTACCTTACTGTGCTTCTTCAAAACATAGACCGTCACATCACAGACGGA
CCATGTCTGATGGAAATCCGACCCCAACTGGTGCAACTATTATCTCAGCCACTGGAGCCTCTGCACTGCCACTCTGCCCCTCACCTGCTCCTCACAGTCATCT
GCCAAGGGAGGTCTCACCCAAGAAGCACAGCACTGTCCACATCGTGCCTCAGCGTCGCCCTGCCTCCCTGAGAAGCCGCTCAGATCTGCCTCAGGCTTACCC
ACAGACAGCAGTGTCTCAGCTGGCACAGACTGCCTGTGTAGTGGGTCGCCCAGGACCACATCCCACCCAATTCCTCGCTGCCAAGGAGAGAACTAAATCCC
ATGTGCCTTCATTACTGGATGCTGACGTGGAAGGTCAGAGCAGGGACTACACTGTGCCACTGTGCAGAATGAGGAGCAAAACCAGCCGGCCATCTATATAT
GAACTGGAGAAAGAATTCCTGTCTTAAACTAAGTGCCTTACTGTTGTTTAAGCATTTTTTTAAGGTGAACAAATGAACACAATGTATCTACCTTTGAACTGTT
TCATGCTGCTGTGTTTTCAAAAGCTGTGGCCATGTTCCTAAATTAGTAAGATATATCCAGCTTCTCAAAAAATGTATATGATTGCTGTTAGCCATGTCTATTG
TTTTTCCTCTGGATTCTTTTCTTATAACTTGGAATACACAAAAGTATAAAACAAGAGATGTGCACCAATGAAAACTATGCTGGGTCGAATTACCTTCAGCACA
ATGTTAATGTTTTCGTTCTCATTTATGCCTTTGTCCATTTGCACACAACAGAAATTGTAATGAGCTTCACTATTTTTGTTTCTTTCCTTCCTTTTTTTTCTTTTTT
CCTTTCTTTCCTTTTTCTTGTCTTGTTTCTTGTTTTTTTCTCTTGTAGTTTCTTTTCTTAATTGTCATTTTTGCAACAAAAAGCCAAGAAAGAGCTTTAGTTTCTT
GGCAAGAATAATGTGATATTAGTAAGTAAAGGTTCTTAAAAGTCTGATGACTGGAATAGATATAAAGTCCTGTTTAAACTACCTAACCTTGGCTGTGGGCCG
ATAATGCATATGTCCAGTTCTCACTTAAATTATGCAATGATATTTCTCTCTGAGGAAATTATACGGAATGTAACTTATAAAAGCTTTACTGAATATAAGTTAT
AAGCATTTTATTCATTAGAACTCCAAAATAGATGTTCAAAGTTCAGTCCTTGCCATTTGACTGAGACCACATGGTGTGCCCCTTGAGTGAGGCTAATCTTTAG
GTTTTTCCTATAGAAAACATTCTTCCTCCATCAGTAGCCCTTTATTTGATATTCAGAAGTGGAAAGCTTTTTCATTCTCCAGTAGAACTTTTAAAAATTGTTAC
AGATACCTAGCTCTTCACAGATATCATGTACTGTAAACAGTCATGTGTCTTAATTTATTTTCTCTATTTGAGTGCATAATTATCCTAATAATCCCAAAGACAC

TGACAACTCAAGGAACAGCAGTACAGTACTATTAGAAGTTAAGTATGTTGTTGTTATTCACATTTCATTTAATTGTGGATAAATGTTAGACATCTGTTGAAA
TAAGCTCATATGGTGGAAAGGACAACTATATTATTGAATTATTTTCAGAAATGGATCTTTGAATAGCAGATCAGGATTTAAATAATAAAATTATCTATGAATC
ACTTTTATGGTCATCATATATGATACAAATCCAGAGTTATTGGTGCAGAAATGGCTACCCGAGAGCTTGGTAAATTTGC

>KSR_H_SEQID#NA_53
GGGAGGGGTGGAAGCGGCAGACTCAGCGGCTCTACCGGCGTCCGGCTCGGGCAGCGGCGCCGAGGGGCGCTCCTGGTCTCCGGCTCGGGGGTT
CCTTGCCGAGGCGCCCGCGGCCGCTCCCAGCCTCGGCGCGCGGCCCGAGCGCTGCGCGGCGGCATGCGAGCAGCGGCGGATGGGAGA
GAAGAAGGAGGGCGGTGGCGGGGATGCGGCGGCGCGGAGGGAGGGCCGCGGCGCAGTGCGGGGCAGCTCCAGAA
GCTCATCGACATCTCCATCGCCATTGTAAGCAGGAGCAGTGCAAGTGCGCTGTGTCCAAGTGACCTCACCCAGGAGATACGAGACCCTGGAGGCGAAGC
TGGTCCGTTACATTTGTAAGCAGGCGGGAGGTGGTGCAGGAGATCCCCGAGACCTCACGTGGCTGTCCCGGTGAGAGACCTGAGATGGTGAAGGAGACGCT
TACACTTTCAAGCTGAGGCCCAGCGGGGGATGAGTTGGCCGTTCAGTATGCCCTGCCAGTCAGGCTCAGGGGAGGACACAAGGAGGAC
GCGGGCGTGTGGGCCAGCGGGGATGAGTTCATTGGAGTGCGGCGGGAAAGTGGCTCAGGAGTCCAGGGCCCCAGGGGAGCTCCCA
TCCAGTTGGAGTTCATTGGATGGCGGGCGGGAAAGTGGCTCAGGAGGCCCACCCTCAGCGACTCCCGCCTCACCCCAGCTTCCAGTGAGGGCCTCTC
GCTGGGCAGAGCAGGCAACAGCGCCAGAGCGCCAGAGCCCCACACCCACCACCACACCCGGAGCAAGTCCCATCACCCGGCCGCCACCAA
AGACAACCTGTATTCCCCTGCACCGACGCGCCCACCCCACCACCCCCGCTCTTCACCCGGTCTGCCCACTCACCGAGCTCTCAGCT
GCTGAAGCCACCACGGACGCCCCCACCCCACCAACGGCCAAGGTCTTCACCGGGAGCAAGTCCCATCACCGGGGATATCGGGCTGTCGGTGACGCACAGGTT
GGGGAACCGCATTGATGAGGTCTCCTGCAGGGTTGATCTCCGATGAGTTGCCCCACCAGATGGTACGGAGGGGATATCGGGCTGTCGGTGACGCACAGGTT
CTCCACCAAGTCCTGGCTGTCGCAGTCTGCGCCACGTGTCGCCACTAACTCGGCTTGGCCTGTCCCTCGGGACATCAACAACCGGTGACAACAAAT
GTACCAAAGAAGAGACCCCTGGTGTGCCCTGTAGAATATCTTCCTGCCACTAACTCGGCTTGGAGGACAGAATCTGTCCCTGGACTCAGCAGAATCAACAACCGGTGACAGAG
CAGCCGAACCCCATTTTGAAACCCTCCCACAACGCCCTCACCGGCGCCCTTCATCATAGACAGAGTTATCTTTCCAGACATTTCAGCCTTTGCACACGCCCGTCCACC
ACCTCCTCCACACCCTCTCCACCGGCGCCCTTCATCATAGACAGAGTTATCTTTCCAGACATTTCAGCCTTTGCACACGCCCGTCCACCCCAACCCTGCCAGGCAGG
TTCAACTTCCCAGCTGCCTACTTCATTCATAGACAGCAGTTTATCTTTCCAGAGCTCACGAAGCTCAGGAGCCAGCTGGCAAGTCAGAGGCAGAAGACGATGA
GTACCCGGCTCGATGACCAGCCAGACTTGCCGAGCTCTGCGGCCGGGGCCCCATCTCTGGGCGGGCCGCAGGGCCCATCGGCAAGTCGTAACCTGCCTGCTCGGAG
GGACGAGGTGGACGACTTGCCGAGCTGAGCGAGCTGGGCGAGCCCACCTGGGGCGGGCCGCAGGGCCGGGCCATGGTGCGGAGGGCCATCGGCGACGACGTCGTTCATGGGGCCTG
CCTTCGAGCAGGTAGAGCTGGGCGAGCCCACCTGGGGCGGGCCGCAGGGCTCTTCAAGAAGAGGTGATGAACTACCGGCACTCGTTTGTGAGGGGACCCCAGAAGGAC
ATGGACGGCCACAACCAGGACCACCTGGCCATTATCACCGGCCATTATCAAGGGCGGGCCGCATGGTACACAAGATCTCAAATCTCAAAGTGTCTTCTATGACAACG
CATGAACCGCCCCACCTGGCCATTATCACCGGGATATCTCATGGGCGGGCCGCTGGTTTGGGATCTCAAGGCGGTGTTTGGGATCTCAAGGCTAAGCGTGTCCCACGACTGGCTGTGC
GAGGCAAATGCGCTCAGGAGATCATCAAGGGCATGGTGGGGATCTCAAGGCGGTGTTTGGGATCTCAAGGCTAAGCGTGTCCCACGACTGGCTGTGC
TATCTGGCCCCTGAGATTGTTACGGCAGGAGTCAGTGGCCCCTGAAGAACCAGGCTGGTGATGTCTATGCATTGGGACTGTTTGG
TATGAGCTGCAAGCTGAGATTGTTACGGCAGGAGTCAGTGGCCCCTGAAGAACCAGGAGATTGGAAGCGGGGGAAGGAATGAAGCGTGTCCTGACTTC
TGTCAGCTTGGGAGGAAGTCAGTGAGATCCTGTCGCGGCCTTTGAAGTCAGGTCACTTAACAGCAGCAAAGTTGTACCCGGTTTGAAAGGTTTTGGCT
AACTTCCCAAGCTGAAGCTGAAGCTGGGAGTCGCGCGCTCTCCAAAGATGTAGCCAGCCATATGTTTTCTTTTTCTTTAAAATGTGTTTCTGAAACATCCCAACAAC
TGGGCCGTCCTGGAGTCGACAAAAAAACACTGCTTAAATCCAGCGCTGCAAAACCAGGAGCACAGTCGTTGGCCATAAACCCACTCGGGAGATTGGAGCTG
CACCACGACAAAAAAACACTGCTTAAATCCAGCGCTGCAAAACCAGGAGCACAGTCGTTGGCCATAAACCCACTCGGGAGATTGGAGCTG
CACTGCTATTTCTTAAAATGAGCAAACAGTCGTGCAGACAACAACCCAACCGTCACGACAGCAGCAAATGTTTACACGTATATTTCTCTGAGTGAACCTGATGTTTTACA
ATAGTAATAAAAAACAGTCTATGCAATTGCAGGTAATTGCAGGTCTTGGGTAGGCCCGTAGGCGGGGCCAGGGGCCCACGGGCGGGTGAGGGACC
CCATGCTGGGCCAGAAGGGAAGGAAGAAGACAAACATGGTAATTGCAGGGTAGGTCTTGGGTAGGCGGGGCCAGGGGCCCACGGGCGGGTGAGGGACC
AGACTCCATCAGGGAGAAATCTATCTAGGGCTCTCCCTTGTCCCTTTCAAAGGATACTGCCCCTTCTCGTCTTGCAGAGGAACCCTGGCTGCAGAACTGAGC

TAGTTTATGGAGTCTGGAATTCCTGGAGAGCTTGGGTTCACCTTCTCACCCCTGTAATCCAGGCTGCTCCTGCTGGAAAAGTAGAAACAGAATCCAAAAAAG
GTCTGGACTCACCCGGTGGTTCCCAGGGTTTCTGCTGCAAGGTGAGGAAACATCCAGTGCTTGTACAGATGTGAGTCTTTGATGAAGCCCCAGGCAG
GCACCAAGGTGATGGGACTCAGGCCTTGGCTTTTAGATACATCCCAGTCACATCTGACACTCGGGGCTGGATGGTGGGAACAAGGAGGAGT
AGATGGCAAAAGTAACCTGAGCCCACTTCCCAGCCACTTCCTGAACCCTTGTCAGTCTGGGCAGTGTGTGGCCATCAGAAGGTTCTAGAACTGCCCACCT
CTTCCATTTCAGTCCTGCTGAAACCCCTTAGCACCCTTAGCAGAGCCTTCCTTCTGTCTGAGTTCGTGTCCATTGTGTCCATTTGTGGCCATCACCCCTTCATTTAG
ACCTACCTAGCTGGCCCCATCTGCAGAGCGCTGTCTTGTCTTTCGTTTAGAAACTCTTAACTGCAGAAAAAGTTCCAGATGGCAAGGGAGCCCTTAAGTGGAGATTAGGTTGCATTAGAC
AAGTGAGAGCACCACGCTTGTCTTCGTTTAGAAACTCTTAACTGCAGAAAAGTTCCAGATGGCAAGGGAGCCCTTAAGTGGAGATTAGGTTGCATTAGAC
TCCAAAACCAGAAAGGAAAAGGGTGATGGGAGACGTGATTGGATTCAGGCCCAGAACCTGTGACCATGCTCTGAGCTCAGACTTGGGAGGGAG
GGGTGTGGCTCCCACCCCTTCCACCCTTCCAGTTAAGGGGTGATGGGAGACGTGATTGGATTCAGGCCCAGAACCTGTGACCATGCTCTGAGCTCAGACTTGGGAGGGAG
AGCTGCTCTGCCAGCAAGCTGTGGGAGCTGTGTGGGAGGCTGTAAACCCGTGGAGTTTCTGTCTATGGTGGGACTGCTCA
CCTGGCCCATCAGCATCTTCACCCTTCGGCCTTCCCACACACCTGTGTGGGGGTCCCCTTGTCAGGGATCACGTGTCTTCATTGCAATGGTGGCCTTTGTCCA
TTTTGCCCCATCATCCCTTTGGCCTTCCCACACACCTGAGCCCCACAGAAAGTCAGTTGGGCATCACGTGTCTTCATTGCAATGGTGGCCTTTGTCCA
GGCAAGAGCCAGGCCGCAGTGGCCATGGGCATGCAAGTGCTGGGCACTGGAACCAGAAAGTCAGTTGGGCATCACAGAACGCTCATTGAGCAAC
CTGACACCAGGCCAGTGGCCATGGGCATGCAAGTGCTGGGCACTGGAACCAGGCCCCACAGAAAGTCAGTTGGGCACTGTGGGGCCCTGTTGTCT
TACAGTGCACTTGGTCTTCTGCAAGTGCTGGGCCCCTCTGCAAGGTGAACTGGGGCACTGGAACTAGAGAACAGTCATGGGGCCCTGTTGTCT
TCAAGGAACCCAGAAGCCACTGGGCCCCTCTGTTCCCAAAGGTGAACTGGGCCCCTCTGTGTCCCTGCCTTAAAGAAC
CCCACCCCACCCCAGATGGCCCCATGATGGCCCCTCTGTTCCCAAAGGTGAACTGAGGGCATGTGGGGGATGGAGAAGGCTCTTAGGGCTGCGGAAG
GAAGGATGACAGATACTGTGAGAGTGGTCTGAGAGTGGGTTTCTGTTCTAACTGCTAGTTGGTGTGTTGAAGTTGATCTCTCCATGTCACCAAATCAA
TCCCACGGGGTCTGAGAGTGGGTTTCTGTTCTAACTGCTAGTTGGTGTGTTGAAGTTGATCTCTCCATGTCACCAAATCAA
GCAGCCCCCTGGGTTTCTGTTCTAACTGCTAGTTGGTGTGTTGAAGTTGATCTCTCCATGTCACCAAATCAA
AACACCCTGTCATCCTAGGGCATTTCCCTTGAATCTTGAGGTCATGTTGTCTATGCAGTGTTTGTACAGTAAAGTTTTATATTTTCATCAACTACATTTGTCTTCCAGACATGCTATTAATTAAATTA
GGTGGAGACGTAAATGTGAAGCCAGTTGGAGTTTGTGTCTATGCAGTGTTTGTACAGTAAAGTTTTATATTTTCATCAACTACATTTGTCTTCCAGACATGCTATTAATTAAATTA
CACACACTTGCTTTTTTGAATGTGATGTAAAATTTGTACAGTAAAGTTTTATATTTTCATCAACTACATTTGTCTTCCAGACATGCTATTAATTAAATTA
AAATGTTAGTATTAACAAACATGCTGTATCGTGATGGTTTTTTTGCCACTGCAAGAACATGCCCCTGTTGTCTAGCCAAGCTGGGTGTCTGGGAGTTGTTGTGAAT
AAAGTTTATACCAAGGTG

>KSR2_H_SEQID#NA_54
ATGGGAGCTGCAGGGCAGAGATGCAGGCCTTCCTCTCTGACTAGGGCACTAGGGCACAGCTGTGGCTGCAGGAAAGGAGGGAGGGGAAGGGGGTTGCAGAGCAAGCGGATGG
ATGAGGAAAACATGACGAAAGGGTCAGGAGCAGCAGAGCCTTGAGTTGCAAAAAGCCTTACAGCAGTGCAAAACATGGTCCAAAACATGATAGACTTGAGCAT
CTCCAACCTGGAAGGGCTTAGGACCAAATGTGCTACCTCCAAACGACCTCACACAAAGAAATCCGGAAGCCCTGGAGAGCAAGCTGGTGAAGTACTTCAGCC
GGCAGCTGTCCTGCAAAAGAAGGTAGCCTTGCAGGAGCGCAAACGGCTTGCAACGGGATGTGGAGGACCTCTTGGAGATGACGGACGATGAAACTCGTGATGTGCGC
AAGGAGGTCCTGGAGGAGTGTGCCGGGATAGTCTCCCGCTCCAACGGCTCTCCTCTCGGTGCCCCCGGAGCCCCCGGAGTGTGCCGGGATCCCCAGAGCCCCAGGGTCCC
CCAGCCACCACAGAGACGGGGAAGGGAAGAGAACAATCCCCACTCCGCCCCTGTGCCCGGGCCCCTTACGGTGCCCGGGATCCGGACAGCTTACGGTGACGTGACCCCCACACCCGCACCCGCACCGTG
AAGTGGTCCAGCACTATTGTCACACCGTCGGGCACGCGCCCCATGGGCCGCCAACAGGCTTACGGTGACGTGACCCCCACACCCGCACCGTGTGCCCG
ACCCGCGCGCTGCATCGGAGCAGCAAATCCCACGGCCCTCCCTGTGAGGCGTGGGGCACCGCGTGGGACGAGCGCCAAAGCCAAGCCCTTGAAC
CACCGCGGGCTGCATCGGAGCAGCGTAGGCAGCGTGCCGAGAACATCCCTCTCAGCAGCGTGCCGAGAACATCCCTCTCAGCAGCCTGGGACATT
CTCAAGATCCACACAGCAGCGTAGGCAGCGTGCCGAGAACATCCCTCTCAGCAGCCTCCGCTCCCGTCGTCCCGGCGCTCCCTCTTCTTTGTGGGACAC

>KIAA1646_H SEQID#NA_55

```
GGATGCAGCTGCTCACTTGGGGCACTGGCCTCTTAGGTTTAAACGATGTCAACAGTGTAGTTGAGAAAATGGCCCGTTAGTGGCTCTATTGCAATAATGTT
AGGGACATTATATGATTTCCACGCAGGTCACACAGTCAGCCTGGGCCTGAGGTAGCAGTGGGTCACTTGCAGGACTTTGATCCACTTTGTAAACGGTTTGTGG
CCAAGTTTTGGGAAGTGGTTGATTCTCTTTGCCTTCAATTCACCTTCATTTGCTTGAGGACATGCTGTGGGGCCTCCACACCTGCTCATCACGGCGGGAG
ATGCAACGTGGCCCTGCTTCAGTGATCCGCGGGAGGGGCCTCCAGCGCCAGACAGATGATAAGCTCTGGAAGCTGTGAAAGCACTGGCAAGTTTGTT
GCTACGACAAATCCGGCTGGGAGCATGCCTTGCGTCTGTTCTGTGGGAGCACAGAGTGATAAGCTCTGGAAGCTGTGGGGGCCATCCGTGGAGGAG
ACTGTTAAAATGTCAAAATACCAAATGCTTTATATCGACGCGAAGTGCTTAACGCGGGCCTGAAGTCTTCCTGTCTTTCTTCACACTGAGTTCTCTAT
GGGCCGGTCCTGACTCCCGCAGGACTCCCTGATGCAGGACTCCCTGTTTACACGTGGTCGTCGTGTTTGTTAGTTGTTAGTTGTTAGTGTGTA
ATTTATTGAACATCTTGTCTTTGAAGCCAGGTAGTGTAAACTGCGTCTGTCTTTGCCTGGGCCACCCTCAATTGTTACTACGTCCTTGTTTAGTGTGTA
GCGCGAGGCGTCCGCGACCAGGAGGATTCTCCCGAAGGCATTCCTGCCGCCGGGCACCCTCAATTGTTACTACGTCCTTGTTTAGTGTGTA
TCCGTGCCCACGTAGACACCGTTCCTCCATGTGTCCGGGGTTTTGCAGAAGCTCCCTTAGTTGATTTCATGTTTAACAGGACCAGCATCCATTTCAGAATTTCCTG
AGCACGCTGTAGACACCGTTCCTCCATGTGTCCGGGGTTTTGCAGAAGCTCCCTTAGTTGATTTCATGTTTAACAGGACGCAGCTTTATGTGGTTTTATGGACC
GCATTGATTTATATTTGAAGCATACAGGAAACTCTCGTTCCTGTTTCAATAATGGTTCAATATATGTTAAGAGTAAATTCTGGGTGTAAAGATAAATGCCCTGGCTCTTTCTACCATGTTGCT
ACATTATCAGAGAGCACTGTCAAGCCAAAGATTTAAGACGGGCTGTGTGCTTCCAGACTGCTTCCACAAAGGTTCTTAGCCTTTATTGTCCTGTAAATGTTT
CCTGGTTGGAAAGAAACCAAAGATTTAAGACGGGCTGTGTGCTTCCAGACTGCTTCCACAAAGGTTCTTAGCCTTTATTGTCCTGTAAATGTTT
CACGGTGGGCGGTGGCTCACGGTGTAATCCAGCACTTGGAGCCGAGGGTGATCACCTAAGGTTCAGGAGTTTGAGATCAGCCTGCCAAC
ATGGTGAAACCTCGTCTACTAAAAATACACAACTTAGCCAGGTCTTGTTGGGCACTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCT
TGAACCCAAGAAGAGAGGGGTGGAGTTGCGGTGAGCCAAGATTGCCGCCACTGCACTCCAGCCTGGGCAAACAGAGGGAGACTCCATCGCCCCCCCAACAA
AAAAAAAAAGTTTCCCATGGCTGTGGTTCCTGAGTGGCTGTTCCTGGTGATACTCTCCAGCAAACAGTCCAGCAAAGATGAGAATCTCAGGACTCGTGTTGCGGTACCG
CCAGCCATTCCGCATGGCCTGTGTTCCTGAGTGGCTGTTCCTGGTGATACTCTCCAGCAAACAGTCCAGCAAAGATGAGAATCTCAGGACTCGTGTTGCGGTACCG
TGTGTCTGACACCTGGCAGCAGCCCTTTGCTATCTGCGCGCAGGAGTTTTAAGGGAAAAAAAAATCAGAAGTGCTGGTTAGATAAAACTTAGGATAAACTTAGGAAAGG
TCAGATGCGGCAGACATACAACACCTGTACTTTAGAGTTTAAGGGAAAAAAAAATCAGAAGTGCTGGTTAGATAAAACTTAGGATAAACTTAGGAAAGG
CTAGTTTTAGCTTCCTTTGTGGCTCCCTGGTGCAAAACAATTAGCAGTTATGCAAACTGATTCTAGTTTATTCTAATTAAGAAAGTGAGGCCGAGTTTGA
CTTCGTTCGTTCCTGAATACAATCTTGAGTAAGTAACTGGAAGTCTGAGTGAAGTCTGAGTGAAGGATGCCTCATTCTCTTTCTAAATCTTGCTGGTTTCAAGATTAGAAAATGGCATT
ATTTGATCTGAAATGTTTGAGAAGACACGAATAAAGTTACTTGGGCAG
```

>DGK-beta_H_SEQID#NA_56

```
TGTGAGCAGAGTTCTTGAAGCTCCACTCCTCTGGGGAAGCCGAGCTGTGTGGGGAGCCGTGTGAATTGGAAAGGATCCTGAGAA
CTGGCTAGTCCCAGTTCCTCTCCGGAAAAGCAGTGGCTCTCGCTTCGCCTGCATCACACTGCATTCTTAACACATTATTGAAA
TTACAAGCATCCAAAGCAGTTCATGTGGACAGATTGCATATTTGAAAGCCTGAGGTATTTTATCATGGAAATCTTTGAGCATAGAC
CTCTGCGCAACACCTGAATAAAGAAATCTTTTACCTGGTATGTGGACAAAGAAATTAAGGGCCCACCTCAGCCCTTC
GGAATTTTCCAACTTGAAAATATGTCGAGTATTCTACAAAGAAATTAAGGATGTCTTGAAGAATTCCATGGTGTGCTTGCAAAGTATAATCC
TGAAGGGAAACAAGACATTCTTAAACAACAAGTTTTTCAAACTTTCAAACTATTCATGAAGACAATTCCTGAAGCTGTTCTTGAGAATGAATAAAGGTGCC
ACACCTTTCATGTCATTTAGCAACAAGTTTCTTCCTCAGTCCAATAACGTGTTCCCGACACCTGAAGGGACTAACCTGAAGAAAATCAGTGCAGAAGAGAAA
ATCACCCTTCCCGAACTACTTCTCTCGACAAATGCTTCCTGACGGATGGGAAGTAATCCATCTGAAGGACAGTCCGAGCTAGAAAATATCATCAGTCAGATGATGTCATG
TTGCAGAATAACCTTGAGTGGGATGGGATGCACTGACTTCACTGAACTTAATCCAATCCTCTGCGGCTTAGAAAATAACGTGAAGAAAATAACGTGGGCGACTGAAGCAC
AATGGATTCAAGGAGGAGAAATGACAACGACCTTTGCCTGAACCTTGCTGCCACTTCACTCTTGTGTCCCTGAACATGCTGAACTTTGCCTGAACATGGGGAAGCAGGCCCTCTGAGTTCCTCTGCGAACATGCTGAACTTGCCTGAACATGGGGAAGCAGGCCCTCTGAGTTCCTCTGCGAACATGCTGGGGAAGCAGGCCCTCTGAGTTCCTCTGCGTTGGCGTGGGGAAGCAGGCGCCCTGTTCCTCTGCGTTGGCGTGGGGAAGCAGGCAGGCCCTCTGTTCCTCTGCAAGTACACAGTCCATGAGCGCT
TTTAACAAACCTGCCTATTGCAACCTTTGCCTGAACATGCTGAACTTGCCTGAACATGGGGAAGCAGGCCCTCTGTTCCTCTGCAAGTACACAGTCCATGAGCGCT
```

GTGTGGCTCGAGCCACCTCCCTCTTGCATCAAGAGACCTATGTGAAGTCCAAAAAGGAACACTGATGTCATGCACCATTACTGGGTTGAAGGTAACTGCCAACCA
AGTGTGATAAGTGCCACCAAAACTGTTAAATGTTACCAGGGCCTGACATCGATCACACTGGTGTCAGTGGTGTCAGATCTGCATAATAAATGTGCTTCTCATC
TAAAACCTGAATGTGACTGTGGACCTTTGAAGGACCATATTTTACCACCACAACAATCTGTCCAGTGTACTGCAGACTCTGCCCACTTCAGGAGTTTCAG
TTCCTGAGGAAAGAACAATCAACAGTGAAAAAGGAAAAGAGTGGTTCCTGGTACTCCTGGTCACTCCTGGTCAGTCTTTCTCTGGAAATGCAAAGAGCCAACTCTGT
TACTGTAGAGATGGACAAGGCCTGCAGGTCACTGTTTACAGTCTTTCTGGAAATGGACCAATGCCAGGGTTAAACTTTTCCGTGATGTTCTGACTTC
TTACAGAAAATTCCAGTATCTATTAAATCCTCGTCAGTCTCGGGCTCGGGTTTGGATTGCATAGAAAAGGCCAAGTAGTGGCAAGCATCCTCCAGTTGCGATTCTGCCTCTT
AGAGTGTTAGCCTGTTGGTGGAGATGGAACCGTGGCTGGATGGGGAGGAGGTTACGAAGGTGCATAGAAGGTGAGAATCTGATGAAAATTCAAAAGACATTGAAAACAGCACAGAAA
GGGACTGGCAATGATCTAGCAGAAGTTTGAAGTTCACACTGAAGAGTCATCATGAGAAGCTACATCGGAGATCCAATCTTTGGGGAGAGTCTAAGAAAGACGTGTCCAATGCAAATGTTGGC
TCATGTTGGACAGGTGGAAGTTTGAAGTTCACACAGATTCCAAGCATGCATGGAGGATCCAATCTTTGGGGAGAGTCTAAGAAAGACGTGTCTCGGTCTCTGTGGAGG
TGGATGCCTCCATTGCACACAGAACTTTCTCAGCCACCTGCAAGAAGCTACATCTTTGGAATATACCGTCACAGAGGCCAAAGAGTTGCAAAGATGGCCAATGA
AATTGCTATTTTGAATATATACCACCGTCACAGAGGCCAAAGAGTTGCAAAGTTGCTGGCCGGCGGCTGAAAAAGTGCTGCCCCATGGAGGATTCAGCCATGCAAATTG
ACAAAAGGACCACCGTCACACAGGCCAAATATACACAGGCTGAAAAAGTGCTGGCCGGCGGCTGAAAAAGTGCTGCCCCATGGAGGATTCAGCCATGCAAATTG
GATGGGCAAATATACACAGGCTGAAAAAGTGCTGGCCGGCGGCTGAAAAAGTGCTGCCCCATGGAGGATTCAGCCATGCAAATTG
ATGGGGAGCCATGGATGCAGAGACCCCATGGAAAACCGAAGGAATAATCCTGTTAGAAATGTTGAAATTGTAAAATCATTAGAAAATT
TCCCTCGTCAAAAGGACAGTCTTTTATAGGCATCTTTGTCTTTCATTTCCTTTTACTAACGTGATTGCATTCCCATAGATTTTTACATC
GAAATCTTTGAAACCATTCAAGTCTTGTATAGGCATCATTTACTAACGTGAGGAAGTTTACCTGAAGTGATACTGAAAAACCGTGAAAACGTGATTCTCAGTGTGATT
TATTGTCTCAGTGTTTTATAGGCATCATTTCCCAATTGCATCAATTCTAACGTGTATATATTGCTATATATAGTGATACTGAAAAACCGTGAAAATA
CCAATTATTACGAAGCACTAATCAGTCAGTAGTTGACTACAACTTGCTATATATAGTGCTATATATAGTGATACTGAAAAACCGTGAAAATA
GGAGAACTTTAATTGACAGTTTTCTGAAAGATTTATTAATTAAAATAGTTGTCTTTTTCTTTTAATTTCTTTTAATTCCAGAAATAGAAGTTTTAAACTA
AAAATATCAGGCATAAGGTTTCCACCAAAGTTCCTTCTGCGAAAAATAGAACGCGGTTTATTGTTCCTTCTGCGTTTTATTTGCTTAGTTTTGTTCATGATGAAAACATGTTACTT
GACAGTCAAGTGTTGGTAATATGCAGTGGTAATATGGCAGCCTGTTCCTTCTGCGAAAAATAGAACGCGGTTTATTGTTCCTTCTGCGTTTTATTTGCTTAGTTTTAAACTA
TGTGCCTAATCAATGTGTTCCACCAAAGTTCCTTCTGCGAAAAATAGAACGCGGTTTATTGTTCCTTCTGCGTTTTATTTGCTTAGTTTTAAACTA
ACAACACAATCAAAGTCAATAGATGATTTCTTTGCTTTTTCAGTTTTCAGTGTAATTATAAATAGGGTACCTGTAATAACTGAGCTCTTGAAGGCGAATCATTGGAGGAGTA
AATGATTGTATCAGTTTATTTCTAGTATAAATGGGTACCTGTAATAACTGAGCTCTTGAAGGCGAATCATTGGAGGAGTA
TCCACTCCATCTTTATGACATTTCTTTTCTTTAGTGTTGTTTAGTGTTCATTGAGGTCTTTAAA
GAAATCACAGAAACAGTACTTTTCTTTTTCTTTTAGTGTTGTTTAGTGTTCATTGAGCAGTCTGTAAATGTTCATTCATTATTGACCTG
CATCTCATCATTGCATTGTTTATGTTTTCAAAACATGCCATAAGGAAACAGAGTGAACTGCATGATTTATTAGTTTCTCTCCACTCTGCATTAAAGTGCT
AATGATTT

>IP6K1_H_SEQID#NA_57
CTTGTTGTTGATCCGTACCCAGTGGGCAGGCGCCGGGACCCAAGCGGCCGGTTGGCTGACCAAGCGGCCGCTGTAGCGGTGCTCAGCGCCACCTGTCGTGCCTGCCAG
GGGGCGGGGCGGGGCGGAAACCTGGAGCGCCCAGCTCCCGTAGGGAGCGCCGTTGGCCGGGGAGCCGTTCGCGTGCGCCCATGTGCTGCATCCTCCTTTGTGCCCATGTGCTGCAACTGGCCCCCCAA
GAATCTGACCATTGGCTTTCACCTGGCCAGGACCTTCACCTGTAGCTCTCCTTTGTGGCCCATGTGCTGCATCCTCCTGGTGCAACTGGCCCCCCAA
CGCAAATGTGTTTGTTCAAACCATGGAAGTGGGCAGTATGCAAGAATGCAAGCCCCTCATCTCCCGGGAGTCCCTCATCTCCCGAAATGA
TAGGCGGGACACAGCAGCATGATGTTACGACGGATCAAGCTGTGCAAGCCCTATGACCTAGAATCAACAGCCTATGACCTAGAATCAACAGCCTAGTTCACTGTTAAA
AGGAGTTCACCCCCTGAATACAAAGGCGTGGTATCGTGGTTACATCAACTTAGTGGCCTATCCTTATGTGGAAAGTGAGA

```
CTGTGGAACAGGATGACACAACAGAAACGGGAGGAGCAACCTCGGCGCGCAAAACACTCCCGGCCGGTCAGGCAGCAGTGGCAGTGACCACAAGGAGGA
GAAAGCCAGCCTGTCCCTGAGACCCTGAGTTCTGAGAAGATCAGGAGTCCAAGAGGTCGGAGCTGCACAGCCACTCAGAGGTCCCTTTCCAGATGCTAG
ATGGCAACAGTGGCTTGAGTTCCTCCTGCTTGAGAACGTGGTGCACCACTCAAGTACCCCTGGCTGCGTTGTCACAAGCAGCTGAGCCGCATGGCCAAGGAC
CGAAAGCTCACAAGTTCCTCCTGCTTGAGAAGTGGGTGGACCTGAAGATGGGCACGCGGCAGCATGGCGATGAC
GCGTCAGCTGAGAAGGCAGCCGGCAGATGCGGAAATGCGAGCAGAGCACATCAGCCACGCGTCAGGGTCTGCGGCATGCAGGTGTACCAGCTGG
ACACAGGGCATTACCTCTGCAGGAACAAGTACTATGGCCGTGGGGCCTGAAAGCTGTGCTGAGATGCCTGAAAGCTGTGCTGAGATGCTCTCAAGCACCTGGACATGTGTCCCTGAGGGTCCCTCTCCTCAGCCACAGGTGCTTCCAGTGGTGTCATCC
TGCGACGTGACCTGTTTGAGCCTATCCTGAGCAACTGCGGGGCCTGAAAGCTGTGCTGAGATGCCTGAAAGCTGTGCTGAGATGCTCTCAAGCACCTGGACATGTGTCCCTGAGGGTCCCTGAGGGTGGCGTCATC
```

>YAB1_H_SEQID#NA_58
GTGGCGGGGCGTCAGCGGCGGTGGCCAGCGGCGGCGGAGAGGCGGCGAGGCGGCTAGAAGGTGACCGGCGGCTCCTGCAGCCAGCCCTGAAGGAT
GGCTGCCATATTGGGAGACACCATCATGTGGCTAAAGGCCTTGTCAAGCTGACCCAGCCGGCCGTTGAAACCCACCTGCGTGGAAGGTGCAGGTGCATGTTCTTGGGGAAGGTGCAGGGTCAGGATAAACATGAAGAATATTTT
AGCTGATCATGGCGGCCAGGGCCTGCAGTCCAGTCCCGCATGCAGCGCGAGCCTCCACAGACTTCTTCTTCAGCCTCCGCTCCGCTCCCGACCAGTCA
GCGCCCCATCCCTGGTCATGCCCACAGCGGAGGCTCTTGCAAACCCAGAGACTCATTCTCTGCCACTGGCTTTCAGCGAAGGTTCTTCCACCAGGACCAATCCCCTGTT
CCTCTGGGCAGGGCCTCACAGCCGAGGAGGCTCTTTGCAAAGGGGACATTGGCGGGCCCGGCCAGGCTAAGGCTCGCCCGAGAACAACAGCAGCACAAACAGCTCAGCGAGCATGCCCCGGAGC
GGAAGGTGCCTGTGACGAGGGATTGGCCGGCTGGCCAACTTCGGAGGTCCGGCCGTGGCCGTGCCAAGCGGATCGTGCGCACGCTCTGCAAGGTGCG
CGCTCCGAGGACCCCTCAGGGAAGAAGAGGCCGCCAGATGCTGGGCCAGATGAGTGGGGCGGG
TGGTGCGGCACTCAAGCTGGGCCAGATGCTGAGCATCCAGGGGACAAGTTGGAATACTTGGAGGGGCCTTGCGGCC
ACTTCATGGCCACTGGAAGCAGCAGATGATGAAAACTCTCAACAAGCGGGCCCCAACTGGCATCAACAGTGATGT
GCATCCATTGGGCAGGTGCACTTGGCCCGAATGAAGCAACATGCTTGAAGGGCCTGTTCCAGAAGGCCTGAAGATCCAGCAACACCTGAGAAGATCCAGTGAGT
CAACAACCTCATGGCCGTGTTGAACATGAGCAACATGTTCCGCGAAGGCCTGTTGAAGGGCCTGTGCCGCAGCC
GTGACTACCAGCGAGAGCCGGAAATATGACAGATCCTTCAGGGACGCCTGCTGCTTCCACTTCATGTGCCCGGCGTTCCTTCATGTGCCCGAGATTGTGGATGGAGCTCGGTTC
CACATGTGCTGACCACAGAGCTGGTTCGAGTTCACTTCAGGGACTCTGCGCCTTCCCCCTGGACCAGCCCAACTGTCAAACAGAACATCTGTTGGATTT
TGTGCCTGAGGGAGCTGTTTCGCACTTCACCGAGATCCTTCACCGGCTCATGAAGAGCTGTTGATGTGCCGCTCGATGTGCGGGCGAAATCCA
TGGGGCAACGCGGGAATATGACAGATCCTCACAGTCAAGGTCACACAGAATCCACACCGTCTGCTTGATGAGCCTT
TAGAGATGAAGGTTCCTCACGGCTACGAGTCAAGGTCAGAAGATCCACACCGGCGGAGAGTGGCCAAGCTTCGTGCGGGCGAAATCCA
TTGATTTTGGCACTCAGAGCCACCACGAGAGATCCAAGCTGGCCAAGGTTCGGCGCAAACTCACTCCTGC
ACAGGAAGATGGGGGCTCCTTCTCATCCTGCTACAGTGGCTGCGGCCCAGGTTCGGGGAGGCTACAGCAACTACTGCAAGAGG
CAGGCCAGCAGTAGGGCTGCGGGCCACGGCTGGGGGTCAGTTCGTGCTAAATGGTTGTAGGGTGAGAAGTGCA
TTTTAACTCCTTTGCCCAATAAGGGGTTTCTGGGTCGGTCAGTCTGCCTCGGTGTCCCTGCGTGTCAGTCTTCCAATCTCTGTTCAGTCAAAACCCAGAGAAACATGAACAGATACGATTGTGGGATTTTATCA
AGAATGAAGATGAAGCCCACTGCTCGGTCAGTCTGCCTCGGTGTCCCTTCACTCTCTGTCGTTTCITTTCCTG
ATGTGAATGTTAAGCACAGAAAGGAAGGAGAGATGTTTCTAGAGTGAAGTGATTGTGAGGATTTTATCA
TCTGTGTAGTAGGTGTGTGTATGTGTTTCTAGAGTGAAGTGATTGTGGGATTTTATCA

>AF052122_H_SEQID#NA_59
GGCAGAGGAGGAAGCCGCCGCTGCCCCCTCAACCCGGGAGGGGGGTCCCCACTGGCTCTGGGCCTCAGGCTCagggtAAtGAtGAGGCGGGGCACAgg
gcACAgggGCTACACTTCAGGTGAGAGTACAGGCAATGTGGCTGAAGGTGGGGACCAGCTGGGCCAGACTGTTGGTTGG
CCTTGTGTGGGGCCTGGGCCCTGGCCCCACCGCGTGGGGACCATGTGGCTTCTTGGGCCCAAAAGTTTTACCAGGATGGCCATCA
GGAGGACATTCGCAGGGCACGGAGGCCCGTGCCCAGGAAGAACGCACCCCAGCTGAGTGACCGCTCTCGAGAAACGGCAAGGTGCCTCCGCAGTCAGAG
GCCGGCTTGGCCAACTTTGGGGACTCAGCCCCTTCCTGTCGCGGAGGCCAATGCGAGCGGATTGTGCAGCGGGGCCGCCCTCAAGGTTGGC
GGTGGTTCTGGGCTTGGGACTCCAGCCCCTTCATCAGCCTGCGAGCGGCATCTTTGCAGGCGGCAGTTCGAGACCTTATGTACAGTTCGAGGGGCCGCCAGATG
CAGATCAGAGTTCTTGAGAGAGGACGGGGACGGGGACCAAGTGGCCTCCTGTGGTGCCTCGCGCTCGCTCAATTGGGCAGGTGCACCA
GGGGCCTGCTGAGGCGGCGGGGACGGGGACGGGCCCTGTTTGCGCAGGCCTTGCAGCGAGAGCAGAGCCCAGAGCATTCAGAGCAGATCCAGGGTACTCA
AGATGAGGCGGCGGGGCCCGGGGCGGGGACCCCTTCTTCCGGGTCCCAGCCGTCGTCGCAAATGACCCCTTCTTCCGGGTCCCAGCGGTGGTTAAGGAGCTGTTGCACGACACGGGTGCTGGGGCATGGGAGCTG
TGTGTGCCCAGGAATTTCAGGCAGCTGCTGCAAATGACCCCTTCTTCCGGGTCCCAGCCGTCGTCGCAAATGACCCCTTCTTCCGGGTCCCAGCGGTGGTTAAGGAGCTGTTGCACGACACGGGTGCTGGGGCATGGGAGCTG

```
GCTGGAGGGGTCCCCTGGACCAGTGCCAGGGCCTAAGCCAGGAGGACCTGCGGAACCAGATTTGCTTCCAGCTCCTGCGACGCTGTGTCTGCGGGAGCTGTTTGAG
TTCCGGATTCATGCAGACTGACCCCAACTGGGCCAACTTCCTGTATGATGCCTCCAGCAGTGACCCTGCTGACTTTGGTGCTCAGGAGTCCTGGGGGAGTTTGGG
ACAGAGTTCACAGACCATTACAATGGAGGTGGTGAAGGCTGCAGCTGATGGAGGCAGTGTGTCCTGCAGAAGACTGTGTCCTGCGAGAAGTCCAGGGACCTCAAATTCCTCACAGG
CTTTGAAACCAAGGCATTCTCCGACGCCACGTGGAGGCAGTGTCGCTGCTGCCACCGGCTTTGCGCACCCGAGGAGACCTATGCCTGCACGTGCCAAGCTGGCAGGGGCTTT
GGCCCGCCGCATACAGGACCTCCGGACCTCATCCGGTGCTGTCGCGCACCGGCTCTTCCAGGACGACCTCTTCCAGGACAGCCTCCATGACAGCCTCCAGACAGCCACTGC
CCTGGCCTGTGCCACCTCCGGACCACCAAAGGGACTCTGGGTGGATCCCTGGGCTTCCTGCGTCTCTCCTGGGGATTCAGTCCCCAGAGAGCAGGCCGTACCTGCTGTAGTGCCTC
CGGCAGCCTCCCACCACCAAAGGGACTCTGGGTGGATCCCTGGGCTTCAGAGGAGCCCCCCTTCCCCCCACTCGAAAGTGGTATCCGCAGGGAAGCGGTTAACTTATCTTTGCCAACATTT
TTATCCCTTCCCGTCTGCCCTGGGTCAGTATCCTGCACCTGCCTAGATCCTGCCACTCCAGAGTCCGGGGTTCGAGATGGTGGCCTG
GGAACTCCCAACTTCGTGCCCGGCTGCATGTCGTTATGCAGATCAGACTCAGGGGGAGCCCTCACCTCTGCGAAAGTGGCGAGATGGTCCTG
GAGGGAGCCTTGGGCGCTGTGCCCTGCTTCTGCTCCCTGCTTCTCCCGGTTCGCGCCCGAGGGCGTCGCGGGGAGCCTCAGTGGTATGCAGTGCAGGCGCCGGCA
AGTTACGGCAGAGAGCCTGCGGCGGAGGAAGCCAGCTACGTGGAGGTAGAGGAGTCCGGGGAGCTCCGGAGTGGCTCGAGTGTGTGAAGAGCTGAAGGCGATGGG
CGTGCAGCTTCCCGGTCAGTGCTCCGGGTACCTGGCTCAGTGGAGGCTCGAGTGGGTCTCGAGTGGTCGGAGTGGTCTGGTGTCGACCATCGAAAAGG
TCTCCTTTTGTGTCCTGACCGGCAAGCTGGGCTTTCTGTGTTCCTATATCTGTCGTGACGGGACTACCCGCCGCTGATCTGCCACTGTTTTCTGCACTGAA
GGACTCCGGCGAGAGGCTGAGCGCCACCAGCTTCGCCCCCACTGTGCCCTGCTGTCTTTGCGCCTGTGTCCCCGACACGGGAGGCCCCGTGCTGTGAGGTCACGGCCGCCT
TCGATGCCAGCGCTGCCCCCCACTGTGCTCGCCTGCTGCCACTGCGCCCATCCCGTGGCTCTGGAGGCCACGATGGGTGAGCGAGCAC
GCAGCAGCTGCCCCACTGCGCCACCACTGCGCGCCATCCCGGCTCGTGGTTCGCGGCTCTTCCAGGGTCCCAGCACTCAG
CCAGAAGAACTCGCCTTTCAAACGGCAGCTGAGCCTACGGCAGCATCAACGCTCTGTGTGCCCCTGCCAGTTCATCTTTTGCACGTGCTGGACGCCAGGGCACAGTGC
CTGAGATGGAGCCTCGTGGTGCGGCGGACAGGGGACTTCTGCCGGGTGAGCCTTGCGCCTGGGCACAAGCGGACCAGCAGCAGCCTCAGAGGCTGAGC
CACCACCTGCCACAACAGGGACTTGTCACAGGTGTCAAGGGCCCTCCGCAGCAGCAGCAGCCCAGTGCCGCGTGTTCCTGCCACCCCCACACATG
GATGGCTGGAGGAGGTGTCACAGTGCCTGCCAGCCTGCAGCCTGGGCTACCCACCGATGCCCCGGCTTTCCCGGCCTTTCCCGGGGCTACCACGCGCAAAGCCCTTC
AGTGCCACCATGCCTCCTGCCTGCCCTACCCGGCTTGGGCTACCCGGGTTGGGCTACGCTGACCGCCGCCCTGTGGGAAGCTGGGGCCTTCCGCCGCATCCCAGCCGCCCCGCCCGC
CAGCCCCCCTTTTGTGCCCCAGTTCAGGCCTACCGGGCTTGGGGCTTCCCGCCCAGTGGCCGCCCTTGAGGCCGGCATTAGAAGGCAAAGC
TGTCTCAGCCGCCCAGTGGCCCCCTGCCCTGAGCCAGGCCAGAGTTGGACCCCTTTGAGGCCGAGCGGCCCCCACCCACTCCATCTCTCCAGGT
CACTGTAGAGAAACCCTCCAAACCCTTTTCTGGGACCTGCAAAAGGCACATTGGAACTGAGATTGAACTCCTAAGGGGCCCCTCCCTCAACCACCCCG
GCCCCACGCCTGGGGGTGGAGGCACAACATTGTGACGGCCCAGGTTGCAACAGGATGGCCACCAGCCAGCTGGCTAAGGGACTTCACTGCGGACTT
ACAACCACTACAGAACCAACATTGTGACGGCCCAGGTTGCAACAGGATGGCCAACAGGGATTGGCACACACTGAATGCCCAGCCAGCTGGTTCACACTCTGTTGACCCC
AGGCTGGCAATGCCCCCCTTCCCCAACCCCAGACACACAGGGGGTTGGCTTAGGGGTGCTTAGGGTCTGCGTTCACTGCGGTCAATATCTGTCGTCTCCTCCCCAT
ACCTTCCAGTGTTGACAGATGGGGCTAAGGGCTAGGGGCCTAGGGCTTCATGGGGCCAGTTGGGGGCCCAGGGCCTAGGGCTGCGTTCACTGCCGTCAATGCTGTTGGGGCCCATGTCTTCCTT
GCACAGCACAAGTCTCATGGGGCTAGGGCCCAGGGCTCAGGGGCCCAGTTGGGGGCCCaGGaggGGGaaGCGGGCCCaGGaGCGGAGGAAGATGGGGAGGAGAAGATGCGCAGTTACCTCATGTCGGTGCCCGCTGG
GGGCCAAGGTCCGGGAAGAAGGGGAAGGGGTGCCTGGCCAGTAATATATGGTGCCCTCGGAAACGAGGGGTGTGTAGGCCCCTGGTGTTTCTATCTTTTGTATCCAAACTTGAAAGATTTGTATTATACAA
GGAGGGGTCCGGGAAGAAGGGGAAGGGGTGCCTGGCCAGTAATATATGGTGCCCTCGGAAACGAGGGGTGTGTAGGCGTAGGCCCCTAGAACCTTTCCTCGAGTCATGGAC
TGAAACAGAGTCAGAAGGATGTCCCTGATGAAGTCCCTGAGTCATAGCGTGGTAACCATAAAAACAAATATTATGAACTGCTGGCCACAGTGCTCACGCCGTGTA
CCAAATTAATCACCCCATTTCTATTCCTTACAGATTAGAATCAAGGAATAAAAACAAATATTATGAACTGCTGGGCACAGTGGCTCTACCAAAAATTC
ATCCCAGCACTTTGGGAGGTCGGAGGATCGAGATCACTTAAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATTC
```

AAAAATTAGCCGGGGCGGTGGTGTGGCAGGCGGCCTGTAATTCCAGGAGGCTGAGGCAGGAGAATGCTTGAACCCGGGCAGCAGAGAGGTTGCAGTGACCTGAGGT
CAACCATTGCACTCCTGCCTGGGCAACAGAATGAGACTCTGTCTCAAAAAAA

>AAF23326_H_SEQID#NA_60
ATGGCCAGAAAGGCTCTCAAGCTTGCTTCGTGGACCAGCAGCATGGCTCTTGCTGCCCTCGGCATCTACTTCTACAGTACAAGTACTTGGACCCTAATGACTTTG
GCGCTGTCAGGGTGGGCAGAGCAGCAGTTGCTACGACGGCTGTCATCAGTTACGACTACCTCACTTCCCTGCCCAGTCCACAAGGCAGTGCTGCATGGATGGGC
AGCTGAGATCTAAGATCCATGATTTGTTCCAGAGCTTCGATGACACCCCTCTGGGGACGGCCCTCCTGCTGATGGAGGTGCTCGTTCTGGCTGTGTGAAGCAGCTGTTC
GGACGGTGGCCGTGAAGGTCCAGCACCCAAAAGTGCGGCGTCAGAGAAGCAAGAAGAAACCTGCCTTTGGAGCTGGAATGCTCAATGGGAGGAAGTGCTGAGAAAGGTGTCCCA
CCAGAGTTTGAGTTTATGTGGCTTGTGTGATGAAGAAGCAAGAGAAACTGCTCCCGAATCCAACTGGGATCTCAATGAGAGGTCCCCGAAATGCTCAGGGCAGGTCAA
GATGCTTCAGGCATTTGACTTTCTTGAAGGTCCCCGAATCCAATGAGAGGTCCACGACCACCTGGGCAAGATGTATAGTGAGATGTCCTGTTGGACCATGGCTTTACCAGATGCTCACG
TGACAGAGACTACATGGAGAGGAACAAGAATGCACCCCGAGACGTCAATGAGATCTCAACGGCACCCCGGCAAAGGGGAGATTGTCCGTGTTGGACCATGGCCAGCGCGGGATCT
GAAGAATTCCGCCTGAATTACTGCCACCTCTGCAGTGGTCGACGGCGGATCGTCGGCAGTCGTCGACGGATCTGGTGCCCTGTGTTGCCTGCATGCTCGATGCTCGGTGGCAT
CTACCCCTTGTTGCCTGCATGCTCGATGCTCTCCCCAGATCAGCCATCTCCTCAACTCCTGCTCAACATGTCCAGCTCGTTGCTCATCAGAGCGCTAGCTGTGCATCAGAGCGCTAGCTGTGCATCGTTC
CAACAACGGCGGCCAACTACCTCCCCGCCCAGCGCCAGCTCCTTTCTCAACTTATGGCGAGGCCTTCAACTTATGGCGAGATCAACCTCCAACTCCATGAGATCAACCTCCATGGCTGGCTGA
ATTCTTCAGAAGGACCCAGATCTCTTCAGCGAGCTCCTGTTCCCTGCTCCACTCTGA
CCGGGTCTTGGCCCTAATATGCTGGCTGGCTGGCTGA

>SGK493_H_SEQID#NA_61
GCTTTGCGGCTGTGCGTCGGGAGAGGCATCTGGGTTCGGACCAGCGGCCGCCGCCATGGGCGAAGTGCGTTACATGAGCCGAGATGACTTCA
GGGTCTTGACCGGCGGTTGAAGAACCATGAAATTGTTCCCGGCAGTTGCTTCTATAGCCAGCCTTAAACATGTGGCTGTGTAATAAAG
TTTTAAGAGAATTAGTGAACAACATAAACTCATAGTCTGGGAGCGTACCAAAACTGTCCAGGCGTACCAAAACTGTCCAGGCGATATGATTACCTAGCTTTGA
AAACACTTTCTTCTAGGCAAGTAGTTGAGTCTGTTGGCAAGAATCAGATATTCACTTGTCAAATGAAGAAGAACAACAA
TTTGCATTAAAGCTTCACAGACTTCACAGACTAGGAAGAATTGCCTTATATGAAGGCATTGTATGAGAGAGAGGAAATTTCAGTTCCAAAACGCGATTATCATAAAACGCACAATGTGTCATGCAGTGGTC
TCTCGTCTCTCTCTGCCATGAAGTGGTTATCCACTATGTCAGATACACCATGTTGAAGATCCTGCATCAGTATATGATGAGCTATGGAACTAATTGTCAAAT
ATGGAACTCATAAAATGGTTATGAGATTTTAATGAATTAATCTCATTTTGGATGAAAGTGACCATATCACCAGGAAAGTGCATCTCAATCTTCAACTTCTCATC
CATGGGCTGATTCATGAGTGTATTTGACAGAGATGTGGAGGTTTCTGCCAGTGAAATCTTTTCCAACTTTAGGATAT
CCAATGCTGAGTGGTATTTGACAGAGATGTGGAGGTTTCTGCCTATTGCAGATCTGAATTTCTCATTAGGTCCAGATGATA
CAGGAGAAGAAGACACTCTTGATGTGGAGGATCTGAATTCTGCCTATTGCGAGTTTATCATCTGAGATCATCGAGAAATGGGCAGAAAATGAAGTGAACGGAA
AAAATATTGAAACAACAAAAGAGGGATCTGAATTCTCATTTTCAGAAATCTGAGATCATCGAGATGTGAAAGATGTGAAAACAGAACTGAA
CTGTCTAGAAATCAGAGGCCTTTCAGAGAGTTCAGAGATTAATCAGAGGAAAACAACTCTGTAAGATGTCAGGCAGGAGAAAAACAGAACTGAA
TTGAAATGACTGAATTCAATCAAGGTCAAGATGGTCAGAGATGGTCAGAGGTCAAGAAGAAAATGTGGGAGCTATGAATCAGTTGACAAAATCAAATCAGAGAATTATATTTAGGATCTTGG
AATAGAGAATTCAGGCTTTGCAGAACTCGAGTTGAACGTCAGTTGACAAAATCAAATCAAATCAAATCAAGTTTGAAAGCTACTGTTTGGGGAGAATAATATTTAGGATCTTGG
GCAAATATATTACCAAGCAACGTAGGGAAAAACATGCAAATATCAAATCAAGTTTGAGCCCTCATTTGTCTTTTGTCTTTTGAGCCAGGCTATCATCATATATATAATAAACCCTCTTTCAT
CT

```
>BRD2_H_SEQID#NA_62
GGACTGCGCGGGATAGGAAGCTGGGGATATGGACAAGCAGCAGCGGCGTTATAGCGCTCTCGGGTTTCGGGACATAGGCCTGGGCCATGCGGCCCCCTTGGCCCCTT
GGCGCGCACCCCCAGGAACGTTCGGAAACGTCGGTCCTGGCTGGGGGAAAGCGGGGAAAGCGGGAAAGCGGGCACGTGGACCCGGTCAGCCAATCTG
GGTGCTGCTGACGTGGCGCGGCGGGGGGGTTGACACCGGATTACATACCCGTCTCCCAGCGTCTCCCAGCCCGTTCCGAAGGCAACTTTGGAGGGCTTAGGATCCAGATTCTCG
TCGTTTTCTGGGGGGGTTACACCCGGATTACATACCCGTCTCCCAGCGTCTCCCAGCCCGTTCCGAAGGCAACTTTGGAGGGCTTAGGATCCAGATTCTCG
CTGCTGCTGCCTTACCGCCGAGAACCACCACCCGCCAGCGTCTTGCGGCCACACCCTGGCCAGTTCCAGTTGGCGTGTCATGGGAAGACTTTGTTGGAAG
CCTGCTTTGGCCAATGGCAGGAGGAGCTGCTGGAGGCTCGGAGGCGATGGCTTCCGCACCTCTTCCCTGGTGTGCCAGTTCCTGAGTCTCCAGTTGGCGTGTCATGGGAAG
GGGAGGCGGGGAGAGAGTGCTGGGGTTGCCACTTCCCGTGTGGGTCTCTGTGGGTGTCTCTGTGGGGCACCTTGGAAATGAAGTTTATGACGTCATCGCTG
TGGCTTGGAGATGTGGCGGGTTGCGGGGTCTCGTGGGTCTCTGACTGAGCTTCTTCACCCGGTGACGAGCGCGCGCGGGAGGGGTGGG
CGGCTGGCCAATAGAAAAGCTCCCGCGGAGAGGTGTTCCTTCCCGCTATATAAAGGGCTGGCGGGGCTCGGCGGCTCGGCGGCCCATTTCGTGTGCTGG
GAAAATCTCAAGCAGGGTGGCGCGCATGAGCCGGCGAAGCTCCTCCTACCGATACAGAGCCCCTGGAAGCCCTAAACCACTTTCGTGTTCATCCGGATCGGGGACCGTCTT
AGTGAGCAGCCTTCGTCCGGCCTATGAGCTGGAGGATTCTGCCTACCGGCAGTACAGAACCCCTGGAAGCCCTAAACCACTTTCGTGTTCATCCGGATCGGGGACCGTCTT
ATGGCCTTCGTCCGGCCTATGAGCTGGAGGATTCTGCCTACCGGCAGTACAGAACCCCTGGAAGCCCTAAACCACTTTCGTGTTCATCCGGATCGGGGACCGTCTT
TGTCCACCCCTCTACTCGGCCCCGTAGGGGCCGACAAGAGAGGAATCCCTGCAGACCAACAGCGGGCTATATTGACGATCGGGATCGGGGACCGTCTTAGCGGGTTATGCTG
GGGACCTCGTCGGCCCCGTAGGGGCCGACAAGAGAGGAATCCCTGCAGACCAACAGCGGGCTATATTGACGATCGGGATCGGGGACCGTCTTAGCGGGTTATGCTG
TTGAAGAGTCAGTCCTCCTTAGTTGCCGGCCTCAGCTGAGCGCAGGGGGGTTTGTGCGGGTTGTGCGGGTCAGAGGATTCGAAAACCCTCTCTCTTGTATGAGGGCT
GAGAGGCCCAAAGACTGCTTTCGTGCGGCCAGCAGGACTTTCCTGCGGTCAAGATGCTGCAAAAGGATTCGAAAACCCTCTCTCTTGTATGAGGGCT
GACCGGGCGGTGAGGGAAACCGAGGCCACCCGAGGGGTTGCTGGGGTTGGGCCCAGAAGCAGCAGCCACCACCCCCGAGGTGTCCAAATCCCAAAAAGCCAGGACGAGTTACCA
TAAGCTCCCTGGGGAGGGAAATGCTTCGGTGCTGCTTTGCAACTTACCCTGTGTGGAAACATCAGTTGCTGGGCCCAGAAGCAGCAGCCACCACCCCCGAGGTGTCCAAATCCCAAAAAGCCAGGACGAGTTACCA
TTGAGAGCCCACAATGCTTCGGTGCTGCTTTGCAACTTACCCTGTGTGGAAACATCAGTTGCTGGGCTCGGCAGCCTGCAAAACTGGGTCTAC
ACCAGCTGCAATATCACAAAATTATAAAACAGCCTATGGGTAGTGATGAAGGCTGCTGTGGAAACATCAGTTGCTGGGCTCGGCAGCCTGCAAAACTGGGTCTAC
CGGATTATCACAAATTATAAAACAGCCTATGGGTAGTGATGAAGGCTGCTGTGGAAACATCAGTTGCTGGGCTCGGCAGCCTGCAAAACTGGGTCTAC
TTAATACCATGTTCACCAAGAGAACAAGAGCTGGTAGTGACATCCTCAAGAGAGGGGCCAAGTTGGCAGCGCCTCCAACATTCCCAGGCAGTGTTACCAGTGC
CAATGCCACAAGAGAACAAGAGCTGGTAGTGACATCCTCAAGAGAGGGGCCAAGTTGGCAGCGCCTCCAACATTCCCAGGCAGTGTTACCAGTGC
CCATCAGGTGGCTTCAAGTCCTTGCACTGTCAGTCCTGGACCCCGGCTCCGGCTCCTGTTACTGCCAGGAGTGCCTCTGGGAGTCTTAGGAAAAGGCGTAAAGCG
CCTCTCCAGTTCTCAAGTCCTTGCACTGTCAGTCCTGGACCCCGGCTCCGGCTCCTGTTACTGCCAGGAGTGCCTCTGGGAGTCTTAGGAAAAGGCGTAAAGCG
GAAAGCAGATACTACCACCCTACACCTACAGCCCCATCAGCCCCACCACGACTTGCGCCCCACCACAGAGTCTAAGGAAAAGGAAAGCTTTCAGAAC
TATGCGTAGAGAGAGTTGCAATGGCATTTGAAGGAGTTACTCTCTAAGAAGACTTGCTCTATGCTGCTCTATGCTGCTCTATAAACCAGTGGATGCACAGGAGTTTGCTGCTG
AGTTAAAACATTGCAATGGCATTTGAAGGAGTTACTCTCTAAGAAGACTTGCTCTATGCTGCTCTATGCTGCTCTATAAACCAGTGGATGCACAGGAGTTTGCTGCTG
GCATGACTACCATGACATCATTAAGCACCTATGGACCTCAGCACTGTCAAGCTCAGCGAAGATGGCAATGGCAAGAAGCTACACAGGAGTGTTGCGTTATG
ATGTACGGCTTATGTTCTCCAACTGCTATAAGTACAATCCCCAGATGTCTTACTGCCATGCCCCTGGAGGAAATTCTCTTGCCAATGTGAAGGTCTCCAGTGAGGAAA
CCAAGATGCCAGATGAAACCAACCACTAGAACCAGGGCCTTTACCAGTTCTACTGCCATGCCCCTGGAGGAAATTCTCTTGCCAATGTGAAGGTCTCCAGTGAGGAAA
GTAGCAGTGAACTACAGAGAGGGGGCAGTTCGGGGCAGTACACTGCTGTCCAATATCCAAGGTCCAATATCCAAGACTCAGAGGAAGAAAAGAGAGAA
ATCGCTTAGCAGAACTACACAGGAAGCAGAACTCGGGGCAGTACAACTGCTGTCCAATATCCAAGACTCAGGTGAAGAATGACAACCTAGGCACCCTCACCC
AAAAGAGAAAAAAGAAAAGCAAGCAAGTGGCAGTGGGGTGCAGAGAAGCAGATCGGAGGAAGCTGAGGAGGCTTAGGCCTCCAAGCTCCCACCTCAACCCCCCAAAA
TAAGAAGTCCAAGAAGAGACAGCCCCACCTGCCTCACCTGCCTCTACAGGTTATGGATTCAGGAGGAAGAGAGCGGCCATGAGTTACGATGAGTTACGATGAGTGAG
AGGCCACAAGACCAAGACAAATTACCTGGGAGAAGCTGGGCCGAGTTGTGCATAATAACTCAAGGCCAGGGAGAAGGAGAGCGGCCATGAGTTACGATGAGTTACGATGAGTGAG
CCTGGGACATCAACAAATTACCTGGGAGAAGCTGGGCCGAGTTGTGCATATAATAACTCAAGCCAGGGAGAAGGAGCGGAGGAGGCCCTTTACGTGATTCAAACCCAGAGAAGAGATTG
```

AGATTGATTTTGAAACACTCAAGCCATCCACCACTTAGAGAGCTTGAGCGGCTATGTCCTTTCCTGCCTACGTAAGAAAACCCGGAAGCCCTACACCATTAAGA
AGCCTGTGGGAAAGAACAAAGGAGGAACTGGCTTTGGAGAAAAAGCGGTTACAAGATGTCAGCGGACAGCTCAATTCTACTAAAAA
GCCCCCAAGAAAGCGAATGAGAAAACAGAGTCATCCTCTGCACAGCAGTAGCCTTAGCGCTTCAGCTCCAGCTCAGACTCCTC
CTCTTCCTCGTCGTCGTCTTCAGACACCAGTGATTCAGACTCAGGGTTGGGGCCAGGAAGGCTCCGGACGGAGACCCTAGACC
ACCCTGCCCACCTGCCCCTTCCCCCTTTGCTGTGACACTTCTTCATCTCACATTGAATTCCCAGCCCCATTGGGAGTGAGTCTTCGCAGTCTCGCAGTGGGGGGATGC
AGGGACATTTACTGAAGGAGGACATGGACAAAACAACATTGAATTCTGAGGCCATAGCTGCCCTATTCACTTCTCTAAGGGCCCTGTTTTGGAGATTGTTGTCTA
CAGGGCAAGGGGTGGGAGTGTGCAAAGCCCTGATCTGGAGTTACCTGAGTGTGGGGCCGTTGGTGGGGGAGGGTGGAAGAAGGGGAGCTCTTTTTTACGTTGATTTT
ATTTATTTTAAGCTAGGTAAGGCTGGGGGAGGTGTTTCCCTTTTCCTTCCGCTCCATTTGGGGCCTCCATTTGGGGCCTCCATTTGGGGGCCGTTCAATCTCCCCATTGTCATCTCCCATTTGGTCCATTGTGTTGATTCTCAAC
TTGTAAAAATAAAGAAAATATT

>BRD3_H_SEQID#NA_63
GACGGCCGCCGGAGCCGCGAGGCCAACTGTCGCCTGGTTGGGCCCGGAAATGGGACGTCGCGCTTTCTCAGGGAGCGTAGAAGCAGCCAGGGCCTCTCCAA
GCGCTGCTGTGACAGAAAGTGAGTGCGGAGGATGTCCACGCCACGAGACAGTGCCCCGGGGATCCCACGAGGACCCGGACCCCGTGAACCCA
CCCCCCCGGAGGTCTCCAACCCCAAGCCCGGCGCCAAGACCAACCAGCTGCAGTACATGTGGTGTGAAGACGCTCTGAAACACCAGTT
CGCCTGGCCCTTCTACCAGCCGCCCGTGAGCGCAATCAAATTGAACCTGCCGGATTATCATAAAAATAATTAAAACCCAATGGATATGGGACTATTAAGAAGA
GACTAGAAAATAATTATTATTGGAGTGCAAGCGAATGTATGCAGGGACTTCAACACCATGTTTACAAATTGTTACATTTATACAAGCCCACAGATGACATAG
TGCTAATGGCCCAAGCTTAGAGAGAAAATTTTCTACAAAAGTGGCCCAAGAGGAAGTTGAATTATTATTACCCCTGTCTCCAAAGGGCAAAGGT
CGGAAGCCGGCTGCGGGAGCCCAGAGCCCGCAGGTACACAGACCAACCATCACCGTGTCTCCCAGCGACCCCTTTCAGAGCGTGCCCCCACCGT
CTCCCAGACGCGTCCCGTTCATCGCTGCCTCCTCCTCCTACGCCGCCCGTTGTCAAACGTCAAGGGCGTGAAGCGGAAAGCAGACACAACCACTCGGCCATCACTGC
CCCATCGTCCCCGTGCTCCCGGCCGTTGTCAGAGACCCGGTTGTCAGACGCGCAGCAGCAAGGTGCAAGCTGTCGAGCCCTGGACTACCAGCAGCAATCAAGCACCGATGGCATCC
CAGCGGAGTGAGTGCACGGGCTACGCTGCCTTCCAGAAGGTGCCCGAGAGTCCAGAAGCTCCAGAGTTTGAGATGAGGTTTGACTCAGGTCTATCCA
AGAAGCACGCACGAGGTTGTGCCATGGCGAGCTGGAGAGAAGGAACACCGATGGAGGAGGCG
CTGCCCCCGCGGGGCCCCCATGGTGAGCCTGGGAGGCTGCAGCTGAAGGCCTGAGGGAGGAAGCACAAAGTGAGGCGGCAGCAGTCGAGGAAGAAGGCGGCAAGAGAGAAG
GAGAAGAAGAAGGAGAAGAAGGAGTCGCAGGAAGAAGGCTCGAGGAAGGCTGGCTC
CGCCTGCCAAGCAGCCTACGACTCAGAGGAAGAAGGCTCCGCCAAGCTTAGCCTGACATCAACGGCTTGCCCGGGGAA
CTGCCTCCTCCTACGACTCAGAGCTAGTGCACAATCTCGGGAACTTGACTTTGAGACTTGAAACCCACCA
GCTGGGCCGGGTAGTGCAACTGGAGAGATGTCCATCTCCAATCTCAATGTCAAGTCTCGTTTACAGAAAAAGCAAAGAACCGTTCTCAGCAAGTCGAAAGA
CTTTGCGGGAACTGCAGGAAAGTCTGGAAAAAGCGGGCCCCGTCGAGGCAGCAGCTCCAGCTCCAGAGTCTGGAGCAGCTCCAGATGTCTGCAGTGACT
CGGCTCAGCACCCTCAGGGGCCCGTCCAGGCTCCTCCAGCAGCAGATGTCGCCACAGCGGTACCCCTCTGTGTTCATTACTCTCTGTTCCATGG
CAGAATGAACTGGCTTCGGACAGGGACAGATGAGGACAGATGGATTCTTTCATATCTTGCTTTCATGGTCAGTGGTTGCTTGAGAGGTGT
TGTGCAGGTCTGCTTCTTAATTCAGTGTTATGATAATCTTCAGGTCAGAGATCTATCTTGCTTAGGTTAGACTTGATGAGAGGTCGT
GAACTCTGCAGAAAGTCTCTTCTTCATCACTGAATTCAGTCACTTGGAGATGCAACTTCAAATGCCCAGAAAACCGTGTGAGATTCGT
ACCGAAGAACCTTGTGGAATCCCTTTGCTTGGTGGAATGTCTCGGATAGCTCGGATAGCTCGGATATGGGTACTGTATTT

GAAAGCTGTTAGCTTGTCAACACGCATTGTCCTTGTCATTGGGCCCGAGCTCTGACCCTCGTGTCTGAGCGCGGCCACCTCTTTCTGGAGGGGCTGAGGAC
AGATGTGCCTGCTCGTGTGGAGACCAGGCTGGGCCTAAGCGGAAGGGTCATCGCAGCGCCCGGAGCGGTGGAGCCCTTGGGGGTGTCGGTGGGATGTG
CGTTCTCCGCTGCTCGTCGTGTGATGTCAGGATCTCCTGGAGGAGCTCGGAGCCTGTGTATGCAGCGGCTGCAGGTTCCATACACTGAGGCTTTTACCTCAACTTT
AAAAAAAAAAGAAGAGAAAAGATTAAAAAAAAAAAAAAAACTCGAGGCATCTATG

>BRD4_H_SEQID#NA_64
ATTCTTTGGAATACTACTGCTAGAAGTCTGACTTAAGACCCAGCTTATGGGCACCAGCTGCTTCTGCAGAGAGAAGGCAGGCCACTGATGGGTAC
AGCAAAGTGTGGTGCTGGCCAAGCCAAAGACCCGTGTAGGATGACTGGGCCTCTGCCCCTTGTGTGGTTGCCACTGTGCTTGAGTGCCTGGTGAAGAA
TGTGATGGGATCACTAGCATGTCTGCGGACATGGCGGCCCTGGGACGCAACGCAGCCAACCTCCAAATCTGCCAGTAAATGGGGGATTGAGAAATCTAC
AACACAGGCCCAGGCCAACCAGCCAGCCAACGCAGCGAGCAACTATGGAAAACACCAGTTTGCATGGCTTTCCAGCAGCCTTTCCGTCAAGCTGAACCTCCT
CAACTGCAATACCTGCTCAGAGTGGTGTCTCAAGACACTATGGGAACAATTAAAGAAGCGCTTGAAAACAACTATTACTGGAATGTATCCAGGACTT
GATTACTATAAGATCATTAAAACGCCTATGGGATGATATGGGAACAAGCCTGGACATAGTCTTAATGGCAGAAGCTCTCTTCTGCAAAAAATAAATGA
CAACACTATGTTTACAAATTGTTACATCTACAAGCCTGGAGATGAGTCATGATCATGTGATAGTCAGGACCCAGCCCTCAGCCCAGCCCTCCTCTGCCGTCACCGGT
GCTACCCACGAAGAAAAACCGAGATCATGATCATGTGATAGTCCAGCAGACCCAGCCCCTCAGCCGCCCTCCTCTGCAGGGTGCCCAGTGCCCAGGAGGAAAAAACAGGACCCTCAGCCCCCC
GGACCTCCATCGTCCAGAACCCCGTGTCATGACAGTCCCGCCTCCGGGTGCCCGGGCTCCAGGGACCCCCCACCCCCCGGCACCACCACCCCCTCCAGC
ACCACCAGCCATTGACCCCATTCACGAGCCCACCCTCGCTGCCCGGAGCCCAGCACCCCAGAGAGCCCCCAAGAGAGCAGCCGCGGCATCCTCTGAAACCTCC
AAAGAAAGGACGTGCCGGACTCTCAGCAGCACCCGCCTACGCCTGGCCCTCTCAGCAGCTCTCAGCAGTGCTGCAGGCAGCAGCGGGGCATCCTCAAGGAGATG
TTTGCCAAGAAGCACGCGCGCCAATCAAGTCTAAACTGGAGGGCCCATGGCCGTGTTTGGCTTTGCAAATGCGTTGAAATGCGTTGTGCTGTGACTACTGTGACATCAAGCACCCCATG
GACATGAGCAGCACAATCAAGTCTAAACTGGAGGTGGTGCCATGGCCCTCCCTCCCCCAAGGTTGTGGCCCGTCGAAGCTCCAAGTGTTCAAACTGTATAAGTAC
AACCCTCCTGACCTGAGGTGGTGCCATGGCCGAGGCGAGCGAGCGCAGCGTGCAAAGCCGTTGACAGTTCGAC
GGCCGGTGTCCCCGGCAGTGAGGCGAGGAGCGAGCGCAGCGTGGAAGAGAAAAGCACAAAGCAAAGAAGAATAAAAAGCAAAGCCAA
TGATGACTTGAGGAGGAGCCAGTGCAGTGAGGCGAGCGCAGCGTTGCAGAGCTCCAGGCCCAGAGA
ACAAACCAAAGAAAAGGAGAAAGAGGAAAAGACACAAGAGAAAAGCACAAAGCAAAGAAGAATAAAAAAGCAAAGCCAA
GGAACCTCTCCTAAAAGACCTAAAAAGAGACGAAGAAAAAAATAATAGCAGCAAGTGTCAGCCTATGTCCTGAAGAAATTCCAACCCCGAGAATTCAACAGCCTATAT
GAGTCGGAGGAAGAGGACAAAGTGCAAGCCCTCCTATAGTGTCCTATGTCCTGAAGAAATTCCAACCCCGAGAGCTCCCACGTATGATTTGAGAGCCGTCCACACTGCGTGCTGAGC
TGGTGCACATCATCCAGTGCACGGGAGCCCTCGTTGTGGAAGAAATCCAACCCGAGAGCTTTGAGACCCGTCCACACTGCGTGCTCCT
TGGAGCGGCTATGTCACCTCCTGTTTGTGCGAAGAAAGGAAACCTCAAGCTGAGAAAGTTGATGTGTCTGCTAATCATTGGACACGGACTCTTAATAAAACGGTCTTC
CAGAGTCGGAGAGCTCCAGGTGAGTCAGGCTCCTGAAGAGAACACTCATTGGACACGGACTCTTAATAAAACGGTCTTC
AGTTCCAGATTCCTTCCCAGCAAGTCTAGCTTAAGTCCATTTCTTCCGTGAAAGGAAGGACCTTATGGAATTCCTCAGAGCCCTGGCCTG
TCCCCCGGGGGTGGATTAGTCATGTCCAGCAGCACACGCCTAGTCCCGCCTTCGGGTGGTCCTGGAGGGTCATCCCAACACACTCCTCTGTAAAGACT
GCCTGGCTGTCCTGCCCAGCCTTCCTGTTCTCTCGTGTTTTCTCTCTCTGGAAGAACACTCTGTGGGACTTGGGACTTCCTAGAGGGCAAGGCCTC
CTACTCGTGTGTCAAAGCCAGAGCCTCAGTGTTCAGTGTCGTTCCCACACCCAAGTATGTTTCTCCATCCTCTTTTGTCCATCTGAGACTCTAGAGCTGTTAGGGACTTGCA
GTGACATCAGTCAGTCCAGGAGCTCAGGAGAGAGCTCTTCCACAGCAGCCCGTGTCCCACAGTCTCAAAGCTTAGAGAGTGGGAGCTTAGAGGGCAAGGCCTC
TAGACCCTGGAAGGCTCAGTGAGGCTCAGTGGTCCCCTGTCCACAGCATGCCTTCTCCACTGGTGCCGTGTAAGCTCGACTCTGAGACTCTGAGTCTTTTCC
TCCTTGCGTCTCCATTGTTCCGTGCATTTCCAAGCTTAAGTTGCTGGTGGGCATTTCCCAGTTTCTATGGGCTCGTCTCTCTCAAGTCACATAGGGAAAGTAC
CTTC

>BRDT_H_SEQID#NA_65
GTGGCAAGATGTTCCTGGGAGGTCAAGTTAAGAGTCAAAAATAATTCATTGAGATTTAACAATTTAGCATGGACATGTACTTGTAGACAGGATTCAAAGCAGT
TAAGAATGTCTCTGCCAAGTCGACAAACAGCTATTATTGTTAAGCTATTATTTATGTGAAGCATAGTTTTCATGGCCCTTTCAACGTCCTGTGGATGCTGTGAAACTAAAGTTGCCTGATTATTA
AGTATCTACAAAAAGTTGTCCTAAAGATTGTTAAATACAATTAAGAAGCGCTTGGAGAATAAAATATTATGCGAAGGCTTCAGAAGTGTATAGAAGACTTCAATACAA
TACCATTATAAAAAAACCAATGGATTTAAATACAATTAAACAAGCCTGGAGATGACATTGTTCTTATGGCACAGAATATAGCTGTTTCTTCTGCTAAAGAAAAATCATCACCCAGCGCAACAG
TGTTCTCAAATTGTTTATTTATATAACAAGCCTGAGATGACAAGGCCACTCAACAGCACATCTATTTCCTAAGACATCTATTTCTCCCTTGAACGTGGTACAGGTGACTTCAGTTCACAAAC
AGAAGAGCAAGTTGTGGGTGTTAAGGAAAGAATTCCTTCTGTTATTCCTAAGACATCTATTTCTCCCTTGAACGTGGTACAGGGAGCTTCAGTGACAAGCAGTTACAAC
AAAAAGTTATTTAAGCAGCAAGAAAGTGTGAAGACTAAAAAGATAAATATGCCAAAGAAATTTTGCCAGATTCTCAGCAACATCCTCCTGCAACTCGAAGTGTTTGCCAGCTGTAAAGTAACT
TGCGGGCCCAAGTTACAAGTTGTGAGCACTGTAGTGCACCTATAAAAGAAATATGCAAAGAAATCTTAAAGGAAATCTTAAGGAAGTATTTTGACGCTGTTAAGTAACT
AAAAATCAGTGGCACTGTAGTGCACCTATAAAAGAAAAGGCACTGTAGTGAGATTCTTAAAGGAAATATTTATCTGTTGTTATAAGTCTGTTGTGACGTTAATGCTTTGG
GAACAAATTAAGGCACTGTAGTGCACCTGTTGTCAAAATCCGATGGATCTTGGAACTATTAAGGAGAAAATGTTGACAAGGATATAAGGATGCATACTCATTTGCG
GCAGATGTTAGATTAATGTTCATGAATTGCTACAAGTACAATCCTCAGATCCAGAGTTGTGACAAGTTGTGACAAGGATGTGACAAGGATGTTTCGAAACGCAT
TTTTCAAAGATCCCGATTGAACCTGTTGAGAGTATGCCTTTATGTTACATCAAAACAGATATCAAAACAGATATCAAACGACTGTTACAGCTCCAGGTTTTG
TCTGAAGGGAACTCTTCTGATTCTGAAGTTCTGATTCTGAAGTATAAAATAAAAAGGAAGAAGTCTAAAAAGGAAAAAGCCAAGAAAACCAAGTTCATTGGTCTAAATCTGAAGATGAAGATAATG
TCCCAAGTACCTTTCCGTAAGCTAAATAAAAGGAAAAGTCCAAGAAAGAAGTCTAAAAAGGAAAAAGCCAAGAAAACCAAGTTCATTGGTCTAAATCTGAAGATGAAGATAATG
ATGTGTGAGCAAATGAGGCTAAGCTATGATGAGAAAAGGCAGTTAGTGAGATAAAACTCCCTGAAGATATAAACAAACTCCCTGAGTAGTTCACATAATACAATCAAGA
CTAAACCTATGGAACTATGATGAGAAAATCCTGATGAGATAGACTTTGAAACTTAAAGAAAAAACAGGAATTGGAAAAGCGGTTACTG
GAGCCTTCTCTGAGCAATTCCAATCTTCTGAGCATCCAACATCTGATGAGATAGACTTTGAAACTTAAAGAAAAAACAGGAATTGGAAAAGCGGTTACTG
TCTAAGAAAGAGACCATTAAAACCTCCTGTAAGAAACGTCAAACATCGATAAAAAATCTGATAAACAAAATCTGATAAACAAATGTTCCGACTGAGTGAGAGCAG
GATGTTAATAATCAGTTAAATTCTAGAAACGTCAAACATCGATAAAAAATCTGATAAACAAAATGTTAAGCTCTTCAGAAATGTTCCTAAGTTTACAG
CAGCAGCAGCAGCTCATCAGAGTTCATCAGAGTGCAGTAGCAGCATGTAAAGAAGAAATGAAGAAGAAACAGGCGTCACACAGATAGGATA
AAGTAAAACCAAATGATTCTCCTTCAGAAGCTCCAAATAATGAAGAAACAGGCGTCACACAGATAGGATA
TTGTGTGCAAGACACAACCTCTGCCAATACTACCCTGTTCATCAGACCAAATTCTGCCTCCCTGAACAGCTCTCAAATGGCATAACTGTGATGCATCAA
GAATTAGAACATTTACAGACTGTGACACAACGATGTTAGAATCTGAAGTCAAGCTCCTGTACAGAAGGATATAAAAGCCATAGAAAGGAAGTTTAGGC
CATCTGGTGATAGTGACACAACGATGTTAATGAAATCCTCAGATGAGTTAATGAAATCCTCAGATGAGTCTTTCAAGCTCTTGGGAACTCGGACACAGG
AAACCAGTGAAACCATCAGTTGTAATGAAATCCTCAGATGAGTCTTTCAAGCTCTTGAAAGCAACTAAAGGTAAAGCTCGGAATCTTTTTCA
AACTCATACGGAAGCATTTGGAACAAAATACAAGGAACATCCCAAATCAGAGGGATCTTGGGAACTAAAGCTCCAAAGTGAGTTATGACTTGGGCGTG
AATAAAATACAAACAAGTGCTCTGGAACAAGAGAGAGGAGGAGGAGAAAGAACATCCCAAGATAAATCCAAGTGACATTATGACTTGGTACCCTTCAAAGTGTTGAAAA
ATTTAGCCAGGCGAAAGAACAAGTGCTCTGGAATTAACCATCACTTAAAATGAATCAAAATGGTAAAAATGATCAAAATGGTAAAAATGATTGCTTTCAGATAACAAGA
CAACTTTGATTAAAAACTCAGTTTTGTATTTGACTGCTCTAAAAATGATTAACCATCACTTAAAATGAATCAAAATGGTAAAAATGATTGCTTTCAGATAACAAGA
TACCAAATCTTATATTGTATTTTGACTGCTCTAAAAATGATTAACCAGTTTTCACTTACAAAAAAAA

>ZC1_H_SEQID#NA_66
CTCGGCGCGCCGCCCACGCTCCGGGCCGTCCTCGAGGCGGCGGCGGCCGGCGCCTGAGGCGGCGCCCGACGCCCGGGGGCC
TGACGGCCGCCCCGCGCCATGGTGTGACGCCGCCGCCTCCGCCCCCGGGCCGGCGCCTCCCGGCGCCGCAGAGAGCCCGAGCGGGCCCGAG
AGCGCAGCCGAGCGCCGCCGCCCCCGGCCGGCGCCGCCTAGCGGCGGCGAGTACCGGGCCGGCGCGGAGGAGGCGGTTCGGCGGCCTGGTCTGCG
GCTGAGATACCACAGAGCGACAGAGAGACACCATTTATTGTTTTTGGTGGTGGCAAAAAGGGAAAAATGGCGAACGACTCCCCTGCAAAAAGTCTGGTGGACA
TCGGACCTCGTCCCTGCGGGATCCTGCTCGGGGATTTTGCTCGGGGATTTTGTTTGTTTTTGTTGGAAATGGCACCTATGGACAAGTCTATAAGGGTTGGACATGTTAAAA

BNSDOCID: <WO____2004006838A2_I_>

```
CGGGTCAGTTGGCAGCCATCAAAGTTATGGATGTCACTGAGGATGAAGAAATCAAACTGGAGATAAATATGCTAAAGAAATACTCTCATCACAGA
AACATTGCAACATATTATGGTGCTTTCATCAAAAAGAGCCTCCAGGACATGATGACCAACTCTGCTTGTTATGGAGTTCTGTGGGCTCTGGTCATTACA
GACCTTGTGAAGAACACCAAAGGGAACACTCAAAGAACTGGATCGCTTACATCTCCAGAGAAATCCTGAGGGGACTGGCACATCTTCACATTCATCA
TGTGATTCACCGGGATATCAAGGGCCAGAATGTGTTGCTGACTGGAAAACTTGTTGACTTTGGTGTGGAGTGCTCAGCTGTGGACAGGACTGT
GGGGCGGAGAAATACGTTCATAGGCACTCCTACTGAGATGGCAGAAGGTGCTCCCCCTCTCGTGACATGCATCCAATGAGAGCACTGTTTCTCATTCCAGAAAACCCTCCT
TTGGTCTTTGTGGCATTACAGCCATTGAGATGGCATTGGAAGAAGTTTTTATAGAAGGGTGCCTGGTGAAGAATTACATGATCGTTGAAGATCATATAGAAGGCAGTTG
CCCGGCTGAAGTCAAAAAATGTCGAAGAAGTTTTTATAGAAGGCAAGTTAGAATCCAGTTAGGATCCAGCTTAAGGATCATATAGAAGGCAGTTG
AAACATCCTTTTATAAGGGATCAGCCAAATGAACAGTTGAGGAGAGAAGTGCCTGAACAGAAGAAGAGGAGCCAAGTTCCATTGTGAACGTGCCTGGTGAGTCTAC
AAACTGAGTATGAGTACAGTTCCTGAGACTGCAGCAGGACTGGTTCCGAGGCTCTTCCGGAGCAGCAACAGTTACTACAGGAGCAACAGCTCCGGGAGCAGG
TCTTCGCCCGAGATTTCCTGAGACTGCAGCAGCGGATTGAGCAGCAGAAGAAACAGAGAAGCGGATTGAGCAGCAGAAAAGAGGCGTCAGAGGCGTCAACAA
AAGAAATATAAAAGGCAACTGCTGGCAGGAACGTGAACAGCGGAAAGAACAGAGGCGTAGAGGAGAGGCAGAGGCGTCAGAGGCAGTTGCAACAA
GAAGCTAGAGACGGGCAGAAGAACTGCTGGCAGGAACGTGAAACAGCGGAGGAACAGAGGAGTATATCAGGTGAAGAGTTGGAGTCCTTCA
GAGGAGACGGGCAGAAGAACTGCTGGCAGGAACGTGAAACAGCGGAGGAGTATATCAGGTGAAGAGTTGGAGTCCTTCA
GCAGCAGCTGCTCCAGGAGCAGGCCATGTTACTGGAGGTGAAAGAGGAGTATATCAGGTGGGGGAGATGCGGCCGCACCGGCGCGCCACCGGA
GAACAAGCACATATCCTGTCTCACAGCCCACTGGAGGCCTGCTGCTCCCGATCAAGCAATGTCTTTTCCCCTGTCTGTTGCATCCGAGTCTCCCGCCCTCTCCAAATTTG
GCTTCCATGTCCCGAGCCCAAAGCCCACTGGAGGCCTGCTGCTCCCGATCAAGCAATGTCTTTTCCCCTGTCTGTTGCATCCGAGTCTCCCGCCCTCTCCAAATTTG
CAGTCTAAGCAGCAGGCAGCAGTATTGGAGCCACCAGTGTTCCTTCCCGATCAAGCAATGTCTTTTCCCCTGTCTGTTGCATCCGAGTCTCCCGCCCTCTCCAAATTTG
CCAGGCAGCCACCAGGTACAGTGTCCCACCTGGTCCCACTGGCATCTCCACCTTCCCCGATCAAGCAGTCTCTTCAGTGAGCTCTCAAGTCAGGAGGCGCCTACCCAAA
CACACCACCATCTCGTTCCAGACCCATCTCCACCTTCCCCGATCAAGCAGTCTCTTCAGTGAGCTGTCTGAGAACAACAGCGCCTGAGATTCCCACTGTCCCAGG
AGGCTTGGTCTAGATCGAGACAGTGACGAGTGCCTCCAAGGGTCCTCCAAGGGTCCTCCACGCCGGTCTCAGAGTTCCTGAGAACAACAGCGCCTGAGATTCCCACTGTCCCAGG
GCAGTGGGCAGCAGAATAGCCAGGCAGGAACCTCAGGGTCCTCCACGCCGGTCTCAGAGTATTGAGCCCGGTCTCACCCTGGTCTCGGGGAACGCTTCAGAGTGAGATCAT
TGGCAGTGGCAGTCCTCCAGGGTCCTCCATCTCAGCGGCTCCAGCAACTCAGAGCTCACCGGGGAACGCTTCAGAGTGAGATCAT
CATCCAAGTCTGAAGGCTCTCCATCTGCAGCCTGGAAAATGCAGCTGGAAAATGCAGTGAAAAAACCTGAAGATAAAAGGAAGTTTCAGACCCCTCCTCAGCCTGCTGGCGAA
GTGGATCTGACCGCACTGGCCAAAGAGCTTGGAGCAGCAGTGAAGATGTGAAGATGTACGGCCACCTCCACAGGACTACTCCTCATCCAGTGAGGAGTCGGGGAC
GACGGATGAGGAGGACGATGTGAAAAAACCATGAGCGGTCGACGAGTGCACCTCAGAGACGCCGGGAGTGCAGCAGCAGTGTCATCTCTGAATTTGAGCAAT
GGTGAAAACGGAATCTGTGAAAAAACCATGATTGTCATGAGAGCAATGGCTTTGCGGGTCGCATCCCAGAGAGCAGCCATCCTAATCGTCCGCCAGAG
TACAGTTGACCAAAGCGTGCCCCATCCTCCAGCCGACCAATGGCTTTGCGGGTCGCATCCCAGAGAGCAGCCATCCTAATCGTCCGCCAGAG
TCCTCCACCTCCTCCTCCATCCTCCAGCCGACCACCATGTCCCCACACGTCCCAGGACAAGCTCACTGTAATGAGACTCAGTCCGCTAGT
AGCACACTCCAGAAACACCAAACAATCTTCCTCCTTTATACACCTTTTATAGAACTCGAGATTGCCTTAGCGGAACAAACAGTGACATCTGTGG
TGGGATTTTCGTGTCGTGATGGGATGAAATACAAGAAGAGGTTTAACTCGAGATTGCCTTAGCGGAACAAACAGTGACATCTGTGG
AGTGACACCCCGGAGATTCGTGCTGGACAGAGTGGCCCAAGGGGAAGGTCTATCTTAACTCGAGATTGCTGCCTTAGCGGTACAGAGAGTGG
CCTGATGCGTCTGGACCAGCAGAGGAGGTGGCTCAAGGGGAAGGTCTATCTTACTCATTATAAAGTTGTAAAATATGAAGAATCAAATTTCTGGTTAAGTGTCTTTGGT
GACAATATGTCTGGCAAAAGGGATAAGTTACGTGTCTACTATTATATAAAGTTGTAAAATATGAAGAATCAAATTTCTGGTTAAGTGTCTTTGGT
GGACAACCGTGGGCACCAAAGCCATATCACAAATTTATGGCCTTAAGTTGTAAATATGAAGAATCAAATTTCTGGTTGAGAAGCAGGGAT
ATGGGTGGGCACCAAAGCCATATCTATGGATCCATCACAAATTTATGGCCTTAAGTCGTGTTGATGTGCTATGAACCCATCAGATGGTGTCTGTGGTGTGTGCTCATGAAGATGAGGGGG
AGAGGTTGAAAGTGATCTATGGATCCATCACAAATTTATGGCCTTAAGTCGTGTTGATGTGCTATGAACCCATCAGATGGTGTCTGTGGTGTGTGCTCATGAAGATGAGGGGG
ACCCACACTCTATGGATCCAGTGAGCATCACAATCAGATGGGGAGAGATGCCTACTACACAGATGTGCTCATGAAGATGAGGGGG
TTTATGTAAACACATATGAAGGATCAGCCAAGGATGTAGTTCTACAGTGAGTTCTACAATATTCGATCCAATCAGACTAACAATGGGGC
TGGGGAGAGAAGGCCATAGAGATCCGGATCGTGGTGGTTCATGCCACTTGGACTGGTGAAACTGGTGTGTTCATGCACAAAGGGCTCAAAAGGGCTCAAAGACTAAAATTCTTGTGTGAACG
```

```
CAATGACAAGGTGTTCTTGCCTCGTCTGTTCGGTCGTCTGGTGGCAGCAGTCAGGTTTATTTCATGACCTTAGGCAGGACTTCTCTGAGCTGGTGAGAAGCAGTGT
GATCCAGGGGATTACTGGCCTCCAGAGTCTTCAAGATCCTGAGAACTTGGAATTCCTGTAACTGGAGCTCGGAGCTGCACCGAGGGCAACCAGGACGACAGCTG
TGTGTGCAGACCTCATGTGTTGGGTTCTCTCCCCTCCTCCGTTCCTCGTTCCTGTCTCAGGTGCTATAAGTGTTGGGATTGAGCATCATACTGGAAAGCAAACACCTTCCTCCAGCTCCAG
GCTGCAATGCAGCTGGTGCTGTTCAGATTCTACCATCAGGTGTCTGACAGTGGCGACGATGAACATGCCGGTTGGTTTATTGCAGTGGCACAAGGAGGTGAGAAGT
AATTCCTTGTCTCTGAATGACTCTGTCTGTCTGTGGGGTGTCTGACAGTGTCTGAAAGCAGAGAGCAGATTAATATGTAAAAATGTCAGTGATAACCTTTTTTGTAGGTGTAAATGTGAAT
GGTGGTAAAAGAAGAGAGAGAGTGCTGAAGCAGAGAGCAGATTAATATGTAAAAATGTGAAAGACACAGTTACTGAGATAACTTTTTTGTGGGTGGGCTTGGGAGATGG
TGGACAAAGAAGAAGAGATGCAGGTTTAAGAAGTTAATATTTATAAAAATGTGAAAGACACAGTTACTGAGATAACTTTTTTGTGGGTGGGCTTGGGAGATGG
GGTGGGGTGGGTTAAGGGGTCCCATTTGTTTCTTTGTTTCTTTGGGATTTGGGGTTGGGATTTGGGATTTGTTGCCTAAATCAT
GTGCAGATGGTTCTAAAAAAAAGAAGAAGAGGTTTCTCAGGTTTAAATTTGTTTAATTTTGTTGTTTAATATGCGTGTGCATTTGTATAATGGCCAGAACTTTGT
CGTGTGACAGTATTAGCACTGCCTCAGTTATTTTTTTTGCCTACCTCAGTTGAATTGAAGTTGCATTTAATTGTCACAAATCAGCAGGCACAAGTTTAGTTATGTTGGGAGAACCATTAATGT
AAAAAATTCAAACCACATGTTTATTTTTTTGCCTACCTCAGTTGAATTGAAGTTGCATTTAATTGTCACAAATCAGCAGGCACAAGAACTGTCATGAAGATGGGAAATAGCCTTA
TTAATTTAATCTTAATAACCAAAACGACCAGAGTTAGAAACGACCAGAGTTAGAAACCAGATTGACTTTTATTGTTTCATAATTGTCACAAATCAGCAGGCACAAGAACTGTCATGAAGATGGGAAATAGCCTTA
AGGCTGATGCAGTTACTTACAAGTTAGAGAACGACCAGAGTTAGAAACCAGATTGACTTTTATTGTTTCATTACCAGATTCACCATTAGAGGTTGATGGGCAACTGCAGCCCATGACACAAGA
TCTCATTGTTCTGATGTAGAGGGTTGGTAGCAGAGGTTGGTAGCAGACAGGTTGGTACAAACTATTTTAAGAATAGTAGTTAAGATGGGCATCTCTGTGTTTAGGTGTAAT
GGGAGTTCCCTTTCTAAAATGCAATGCACTAAACTATTTTAAGAATGTAGTTAAGATGGGCATCTCTGTGTTTAGGTGTAAT
ATCGAAGTCCTGGCTTTCTGGCTTTCTCTGTTTCTCTGTTTTTTAAACCATTTACTTTATGAATATATTCATGACTCCATGTCCAGTCCTGCCA
TGAGAAAGAATTGTTCATGTCTGAGTTGAGAAATAACTGCATTATGGCTAATAACTGCATTATGGCTAAGGAAGGATCAAAGAAACTGCGTTTTTTTAGGCTTGATCTTTT
TCTTACTTTCCTCTGTTGAGTTGAGTTGCAGTTCCTCGTTCCGTTCCGTTCCATCTTTTGAAATAACTGCATTATGGCTGTCCCTGCTTAACCTGGAAATAACTGTGTTTGTTTGCTTCCCATCCC
ACTTGGCACCCACTCTGACATGCAGAGCAGCAGCTGGTTTGTGTCAGTTCCGTTCTGTGCGTGTTTGTTTGTGTATACCTGTAAAGACAAGCTTACTTTTGGGAAGTCAAGACAAGTACAT
CCACTCATGAATGCAGAACACAGCAGCTGGGAGCAGTTGGACAGTTGTGAAGGCCAAAGCCGCCTACTGTGTTTGTGTATACCTGTAAAGACAAGCTTACTTTTGGGAAGTTAAAGAAGATGGAAATGG
ATGTTTCATTTGTATGAGTTGGAGCTGGATTCCATGCCACAGGACTTTAGAATTAAATTCCTGCTTAGTTGTAATTCCTGCTTAGTTGTAATTCCTGGCTTTTCATAAGATGAAACAT
GGATTTGAGGTAGGCTGGATTCCATGCCACAGGACTTTAGAATTAAATTCCTGCTTAGTTGTAATTCCTGTGTAATTCCTGTGTAATTCCTTGTGGGCAGATGTCTTTGAAGAAATGCTCAATG
GCCACAGTGAACACACTCGTGTACATATCAAGAGAACAAAGGATGGCATAAGTGTCTTAACATTATCCAAAGACTTGAAGGTGGGCAGATGTCTTTGAAGAAATGCTCAATG
TCTGATGCTAAGTGGGAGAAGGCAGAGAACAAAGGATGGCATAAGTGTCTTAACATTATCCAAAGACTTGAAGGTGGGCAGACTCCATGGGGTCCATGTCTGTAAGTCAAATGTTAC
ACAAAAAAAATGCAAATGGTGTTCATTGGGAATTACCAAGTGTTCATTGGGAATTACCAAGTGCTTGAGAACTTGCTGGCTTTCCCATAGGGGTCCATGTGGGCTTAGTTACTACCTTTAAAG
GCTTGGCTTGCTGTGCCTTCTCGGCGGGGGCTTCCGGCTTCGTGCTTTACTGGCTTTGTGGTGTGTTCCATTGTTGTGGGCTTAGTTACTACCTTTAAAG
ACACCCTTTGTCCTTCTCGGCGGGGGCTTCCGGCTTCGTGCTTTACTGGCTTTGTCCCGTCTTTTTCCCGTCTTTGTCAACCCAGATTGTTGCTAT
AGCTTGATGCAGGGACCCAGTGAAGTTTCTCCGTTAAAGATTGGGAGTCGTCGAAAAACAGTATTGCAACATTACTTTAGGAAGGAATTATTTCCCCCCTTTTACAGG
GTAGTAACTTCTCCAAACAAACAAAATGTCTCTAGGTTAGTCGTTGAGTCAGAAAACAGTATTGCAACATTACAAGGAACATTGAACCTTTCTGTTAGAGGAGT
GCTCTCCAAACAAACAAAATGTCTCTAGGTTAGTCGTTGAGTCAGAAAATGTTTACAAAATGGTTTGTTACACTTACCTTATTAAATCAGAGTAACCCATATTTATGGTTATTTTAA
AGTGTTTTTTACTCTTTCTCATGCACATGTTAGGTTCGTTCGTTGTTAGCACACTTACCTTATTAAATCAGAGTAACCCATATTTATGGTTATTTTAA
GTGACTTTTATGGTTATTTATGGTTGTTTCGTTCGTTGTTAGCACACTTACCTTATTAAATCAGAGTAACCCATATTTGTGATTATTGAGTTAATAGTAATACAAATGACAAACTGC
ATTAAATTTACTAATTATAAAAGCTGCAAAGCACTGTGGCAAGTCGTGGCAAGCAGCCCTTTTTTTGCAGTGCTAACTTGTCTACTGTTACTGTTATGAAATTACT
GTTGTCCCCCACCCTTTTTCCTTAAATAAAGTAAAAATGACACCTA
```

Figure 2

>CRIK_H_SEQID#AA_67
MLKFKYGARNPLDAGAAEPIASRASRLNLFFQGKPPFMTQQQMSPLSREGILDALFVLFEECSQPALMKIKHVSNFVPEVYSDTIAELQELQPSAKDFEVRSLVGCG
HFAEVQVVREKATGDIYAMKVMKKALLAQEQVSFFEEERNLSRSTSPWPLQYAFQDKNHLYLVMEYQPGGDLLSLLNRYEDQLDENLIQFYLAELILAVHS
VHLMGYVHRDIKPENLVDRTGHIKLVDFGSAAKMNSNKMVNAKLPIGTPDYMAPEVLTVMNGDKGTYGLDCDWWSVGVIAYEMIYGRSPFAEGTSARTFNNI
MNFQRFLKFPDDPKVSSDFLDLIQSLLCGQKERLKFEGLCCHPFFSKIDWNNIRNSPPPFVPTLKSDDDTSNFDEPEKNSWVSSSPCQLSPSGFSGEELPFVGFSYSKA
LGILGRSESVVSGLDSPAKTSSMEKKLLIKSKELQDSQDKCHKMEQEMTRLHRRVSEVEAVLSQKEVELKASETQRSLLEQDLATYITECSSLKRSLEQARMEVSQE
DDKALQLLHDIREQSRKLQEIKEQEYQAQVEEMRLMMNQLEEDLVSARRRSDLYESELRESRLAAEEFKRKATECQHKLLKAKDQGKPEVGEYAKLEKINAEQQ
LKIQELQEKLEKAVKASTEATELLQNIRQAKERAERELEKLQNREDSSEGIRKKLVEAEERRHSLENKVKRLETMERRENRLKDDIQTKSQQIQQMADKILELEEKH
REAQVSAQHLEVHLKQKEQHYEBKIKVLDNQIKKDLADKETLENMMQRHEEEAHEKGKILSEQKAMINAMDSKIRSLEQRIVELSEANKLAANSSLFTQRNMKAQ
EEMISELRQQKFYLETQAGKLEAQNRKLEEQLEKISHQDHSDKNRLLELETRLREVSLEHEEQKLELKRQLTELQLSLQERESQLTALQAARAALESQLRQAKTELE
ETTABEAEEIQALTAHRDBIQRKFDALRNSCTVITDLEEQLNQLTEDNAELNNQNFYLSKQLDEASGANDEIVQLRSEVDHLREITEREMQLTSQKQTMEALKTTC
TMLEEQVMDLEALNDELLEKERQWEAWRSVLGDEKSQFECRVRELQRMLDTEKQSRARADQRITESRQVVELAVKEHKAEILAIQQALKEQALKEPSSLHEGWMKVPR
LEKKHAMLEMNARSLQQKLETERELKQRLLEEQAKLQQQMDLQKNHIFRLTQGLQEALDRADLLKTERSDLEYQLENIQVLYSHEKVKMEGTISQQTKLIDFLQA
KMDQPAKKKKVPLQYNELKLALEKEKARCAELEEALQKTRIELRSAREEAAHRKATDHPHPSTPATARQQIAMSAIVRSPEHQPSAMSLLAPPSSRRKESSTPEEFS
RRLKERMHHNIPHRFNVGLNMRATKCAVCLDTVHFGRQASKCLECQVMCHPKCSTCLPATCGLPAEYAIHFTEAFCRDKMNSPGLQTKEPSSSLHEGWMKVPR
NNKRGQQGWDRKYITVLEGSKVLIYDNEAREAGQRPVEEFELCLPDGDVSHGAVGASELANTAKADVPYILKMESHPHTTCWPGRTLYLLAPSFPDKQRWVTALE
SVVAGGRVSREKAEADAKLLGNSLLKLEGDDRLDMNCTLPFSDQVVLGTEEGLYALNVLKNSLTHVPGIGAVFQIYIIKDLEKLLMIAGEBERALCLVDVKKVKQ
SLAQSHLPAQPDISPNIFEAVKGCHLFGAGKIENGLCICAAMPSKVVILRYNENLSKYCIRKEIETSEPCSCIHFTNYSILIGTNKFYEIDMKQYTLEEFLDKNDHSLAP
AVFAASSNSFPVSIVQVNSAGQREEYLLCFHEFGVFVDSYGRRSRTDDLKWSRLPLAFAYREPYLFVTHFNSLEVIEIQARSSAGTPARAYLDIPNPRYLGPAISSGAI
YLASSYQDKLRVICCKGNLVKESGTEHHRGPSTSRSSPNKRGPPTYNEHITKRVASSPAPPEGPSHPREPSTPHRYREGRTELRRDKSPGRPLEREKSPGRMLSTRRE
RSPGRLFEDSSRGRLPAGAVRTPLSQVNKVWDQSSV

>DMPK2_H_SEQID#AA_68
MERRLRALEQLARGBAGGCPGLDGLLDLLLALHHELSSGPLRRERSVAQFLSWASPFVSKVKELRLQRDDFEILKVIGRGAFGEVTVVRQRDTGQIFAMKMLHKW
EMLKRAETACFRBERDVLVKGDSRWVTTLHYAFQDEEYLYLVMDYYAGGDLLTLLSRFEDRLPPELAQFYLAEMVLAIHSLHQLGYVHRDVKPDNVLLDVNGHI
RLADFGSCLRLNTNGMVDSSVAVGTPDYISPEILQAMEEGKGHYGPQCDWWSLGVCAYELLFGETPFYAESLVETYGKIMNHEDHLQFPPDVPDVPASAQDLIRQ
LLCRQEERLGRGGLDDFRNHPFFEGVDWERLASSTAPYIPELRGPMDTSNFDVDDDTLNHPGTLPPPSHGAFSGHHLPFVGFTYTSGSHSPESSSEAWAALERKLQC
LEQEKVELSRKHQEALHAPTDHRELEQLRKEVQTLRDRLPEMLRDKASLSQTDGPPAGSPGQDSDLRQELDRLHRELAEGRAGLQAOEQELCRAQGQQEELLQRL
QEAQEREAATASQTRALSSQLEEARAAQRELEAQVSSLSRQVTQLQGQWEQRLEESSQAKTIHTASETNGMGPPEGGPQEAQLRKEVAALREQLEQAHSHRPSGK
BEALCQLQEENRRLSREQERLEAELAQEQESKQRLEGERRETESNWEAQLADILSWVNDEKVSRGYLQALATKMAEELESLRNVGTQTLPARPLKMEASARLELQ
SALEAEIRAKQGLQERLTQVQEAQLQAERRLQEABKQSOALQQELAMLREELRARGPVDTKPSNSLIPFLSFRSSEKDSAKDPGISGEATRHGGEPDLRPEGRRSLR
MGAVFPRAPTANTASTEGLPAKGWGMGPWEALGNGCPPPQPSHTLRPRSFPSPTKCLRCTSLMLGLGRQGLGCDACGYFCHTTCAPQAPPCPVPPDLLRTALGV
HPETGTGTAYEGFLSVPRPSGVRRGWQRVFAALSDSRLLLFDAPDLRLSPPSGALLQVLDLRDPQFSATPVLASDVIHAQSRDLPRIFRVTTSQLAVPPTTCTVLLA
ESEGERBRWLQVLGELQRLLLDARPRRPRVYTLKEAYDNGLPLLPHTLCAAILDQDRLALGTEEGLFVIHLRSNDIFQVGECRRVQQLTLSPSAGLLVVLCGRGPSV
RLFALAELENIEVAGAKIPESRGCQVLAAGSILQARITPVLCVAVKRQVLCYQLGPGPWQRRIRELQAPATVQSLGLLGDRLCVGAAGFALYPLLNEAAPLALG

AGLVPEELPPSRGGLGEALGAVELSLSEFLLLFTTAGIYVDGAGRKSRGHELLWPAAPMGWGYAAPYLTVFSENSIDVFDVRRAEWVQTVPLKKVRPLNPEGSLFL
YGTEKVRLTYLRNQLAEKDEFDIPDLTDNSRRQLFRTKSKRRFFFRVSEEQQKQQRREMLKDPFVRSKLISPPTNFNHLVHVGPANGRPGARDKSPSQPLRTVTQQ
APEEKGRVARGSGPQRPHSFSEALRRPASMGSEGLGGDADPTGAVKRKPWTSLSSESVSCPQGSLSPATSLMQVSERPRSLPLSPELESSP

>MAST3_H_SEQID#AA_69
MKSRRDKLHIPALTLDLSPSSQSPSLLGPSSPCSPCSPSLGLHPWSCRSGNRKSLVVGTPSPTLSRPLSPLSVPTAGSSPLDSPRNFSAASALNFPFARRADGRRWSLAS
LPSSGYGTNTPSSTLSSSSSRERLHQLPFQPTPDELHFLSKHFRSSENVLDEEGGRSPRLRPRSRSLSPGRATGTFDNEIVMMNHVYRERFPKATAQMBGRLQEFLTA
YAPGARLALADGVLGFIHHQIVELARDCLAKSGENLVTSRYFLEMQEKLERLLQDAHERSDSEEVSFIVQLVRKLLIIISRPARLLECLEFDPEEFYHLLEAAEGHAR
EGQGIKTDLPQYIIGQLGLAKDPLEEMVPLSHLEEEQPPAPESPESRALVGQSRRKPCESDFETIKLISNGAYGAVYLVRHDTRQRFAIKKINKQNLLRNQIQQVFV
ERDLLTFAENPFVVSMFCSFETRRHLCMVMEYVEGGDCATLLKNMGPLPVDMARLYFAETVLALEYLHNYGIVHRDLKPDNLLITSLGHIKLTDFGLSKIGLMSMA
TNLYEGHIEKDAREFIDKQVCGTPEYIAPEVIFRQGYGKPVDWWAMGVVLYEFLVGCVPFFGDTPEELFGQVVSDEIMWPEGDEALPADAQDLITRLLRQSPLDRL
GTGGTHEVKQHPFFLALDWAGLLRHKAEFVPQLEAEDDTSYFDTRSERYRHLGSEDDETNDEESSTEIPQFSSCSHRFSKVYSSSEFLAVQPTPTFABRSFSEDREEG
WERSEVDYGRRLSADRLRSWTSSGSSCQSSSSQPERGPSPSLLNTISLDTMPKFAFSSEDEGVGPGPAGPKRPVFILGEPDPPPAATPVMPKPSSLSADTAALSHARL
RSNSIGARHSTPRPLDAGRGRRLGGPRDPAPEKSRASSSGGSGGGRVPKSASVSALSIITADDGSGGPLMSPLSPRSLSSNPSSRDSPSPVCGSLRPPIVIH
SSGKKYGFSLRAIRVYMGDSDVYTVHHVVWSVEDGSPAQEAGLRAGDLITHINGESVLGLVHMDVVELLLKSGNKISLRTTALENTSKVGPARKNVAKGRMAR
RSKRSRRRETQDRRKSLFKKISKQTSVLHTSRSFSSGLHHSLSSSESLPGSPTHSLSPSPTTPCRSPAPDVPADTTASPPSASPSSSPASPAAAGHTRPSSLHGLAAKLG
PPRPKTGRRKSTSSIPPSPLACPPISAPPPRSPSPLPGHPPAPARSPRLRRGQSADKLGTGERLLDGEAGRRTRGPEAELVVMRRLHLSERRDSFKKQEAVQEVSFDEPQ
EEATGLPTSVPQIAVEGEEAVPVALGPTGRD

>MAST205_H_SEQID#AA_70
MKRSRCRDRPQPPPDRREDGVQRAAELSQSLPPRRRAPPGRQRLEERTGPAGPEGKEQDVVTGVSPLLFRKLSNPDIFSSTGKVKLQRQLSQDDCKLWRGNLASS
LSGKQLLPLSSSVHSSVGQVTWQSSGEASNLVRMRNQSLGQSAPSLTAGLKELSLPRRGSFCRTSNRKSLIVTSSTSPTLPRPHSPLHGHTGNSPLDSPRNFSPNAPAH
FSFVPARRTDGRRWSLASLPSSGYGTNTPSSTVSSSCSSQEKLHQLPFQPTADELHFLTKHFSTESVPDEEGRQSPAMRPRSRSLSPGRSPVSFDSEIMMNHVYKERF
PKATAQMEERLAEFISSNTPDSVLPLADGALSFIHHQVIEMARDCLDKSRSGLITSQYFYELQENLEKLLQDAHERSESSEVAFVMQLVKKLMIIIARPARLLECLEFD
PEEFYHLLEAAEGHAKEGOQIKCDIPRYTVSQLGLTRDPLEEMAQLSSCDSPDTPETDDSIEGHGASLPSKKTPSEEDFETIKLISNGAYGAVFLVRHKSTRQRFAMK
KINKQNLLRNQIQQAFVERDILTFAENPFVVSMFCSFDTKRHLCMVMEYVEGGDCATLLKNIGALPVDMVRLYFAETVLALEYLHNYGIVHRDLKPDNLLITSMG
HIKLTDFGLSKIGLMSLTTNLYEGHIEKDAREFLDKQVCGTPEYIAPEVILRQGYGKPVDWWAMGIILYEFLVGCVPFFGDTPEELFGQVISDEIVWPEGDEALPPDA
QDLISKLLHQNPLERLGTGSAYEVKQHPFFTGLDWTGLLRQKAEFPQLESEDDTSYFDTRSERYHHMDSEDEEVSEDGCLEIRQFSSCSPRFNKVYSSMERLSLL
EERRTPPPTKRSLSEEKEAHSDGLAGLKGRDRSWVIGSPEILRKRLSVSESSHTESDSSPPMTVRRRCSGLLDAPRFPEGPEEASSTLRRQPQEGIWVLTPPSGEGVSG
PVTEHSGEBQRPKLDEEAVGRSSSGSSPAMETRGRGTSQLAEGATAKAISDLAVRRARHRLLSGDSTEKRTARPVNKVIKSASATALSLLIPSEHHTCSPLASPMSPHSQ
SSNPSSRDSSPSRDFLPALGSMRPPIHHRAGKKYGFTLRAIRVYMGDSDVYTVHHMVWHVEDGPASEAGLRQGDLITHVNGEPVHGLVHTEVVBLILKSGNKVAI
STTPLENTSIKVGPARKGSYKAKMARRSKRSRGKDGQESRKRSSLFRKITKQASLLHTSRSLSSLNRSLSSGESGPGSPTHSHSLSPRSPTQGYRVTPDAVHSVGGNS
SQSSSPSSSVPSSPAGSGHTRPSSLHGLAPKLQRQYRSPRRKSAGSIPLSPLAHTPSPPPPTASPQRSPSPLSGHVAQAFPTKLHLSPPLGRQLSRPKSAEPPRSPLLKRV
QSAEKLAAALAASEKKLATSRKHSLDLPHSELKKELPPREVSPLEVVGARSVLSGKGALPGKGVLQPAPSRALGTLRQDRAERRESLQKQEAIREVDSSEDDTEEGP
ENSQGAQELSLAPHPEVSQSVAPKGAGESGEEDPFPSRDPRSLGPMVPSLLTGITLGPPRMESPSGPHRRLGSPQAIEEAASSSAGPNLGQSGATDPIPPEGCWKAQ
HLHTQALTALSPSTSGLTPTSSCSPPSSTSGKLSMWSWKSLIBGPDRASPSRKATMAGGLANLQDLENTTPAQPKNLSPREQGKTQPPSAPRLAHPSYEDPSQGWLW
ESECAQAVKEDPALSITQVPDASGDRRQDVPCRGCPLTQKSEPSLRRGQEPGGHQKHRDLALVPDELLKQT

>MASTL_H_SEQID#AA_71

MDPTAGSKKEPGGGAATEEGVNRIAVPKPPSIEEFSIVKPISRGAFGKVYLGQKGKGKLYAVKVKKADMINKNMTHQVQAERDALALSKSPFIVHLYYSLQSANN
VYLVMEYLIGGDVKSLLHIYGYFDEEMAVKYISEVALALDYLHRHGIIHRDLKPDNMLISNEGHIKLTDFGLSKVTILNRDINMMDILTTPSMAKPRQDYSRTPGQV
LSLISSLGFNTPIAEKNQDPANILSACLSETSQLSQGLVCPMSVDQKDTTPYSSKLLKSCLETVASNPGMPVKCLTSNLLQSRKRLATSSASSQSHTFISSVESECHSSP
KWEKDCQESDEALGPTMMSWNAVEKLCAKSANAIETKGFNKKDLELALSPIHNSSALPTTGRSCVNLAKKCFSGEVSWEAVELDVNNINMDITDTSQLGFHQSNQ
WAVDSGGISEEHLGKRSLKRNFELVDSSPCKKIIQNKKTCVEYKHNEMTNCYTNQNTGLTVEVQDLKLSVHKSQQNDCANKENIVNSFTDKQQTPEKLPIPMIAKN
LMCELDEDCEKNSKRDYLSSSFLCSDDDRASKNISMNSDSSFPGISIMESPLESQPLDSDRSIKESSFEESNEDPLIVTPDCQEKTSPKGVENPAVQESNQKMLGPPLE
VLKTLASKRNAVAFRSFNSHINASNNSEPSRMNMTSLDAMDISCAYSGSYPMAITPTQKRRSCMPHQTPNQIKSGTPYRTPKSVRRGVAPVDDGRILGTPDYLAPEL
LLGRAHGPAVDWWALGVCLFERLTGIPPFNDETPQQVFQNILKRDIPWPEGEEKLSDNAQSAVEILLTDDTKRAGMKELKRHPLFSDVDWENLQHQTMPFIPQPD
DETDTSYFEARNTAQHLTVSGFSL

>PKC_eta_H_SEQID#AA_72
MSSGTMKFNGYLRVRIGEAVGLQPTRWSLRHSLFKKGHQLLDPYLTVSVDQVRVGQTSTKQKTNKPTYNEEFCANVTDGGHLELAVFHETPLGYDHFVANCTLQ
FQELLRTTGASDTFEGWVDLEPEGKVFVVTILTGSFTEATLQRDRIFKHFTRKRQRAMRRRVHQINGHKFMATYLRQPTYCSHCREFIWGVFGKQGYQCQVCTCV
VHKRCHHLIVTACTCQNNINKVDSKIAEQRPGINIPHKFSIHNYKVPTFCDHCGSLLWGIMRQGLQCKICKMNVHRCQANVAPNCGVNAVELAKTLAGMGLQPG
NISPTSKLVSRSTLRRQGKESSKEGNGIGVNSSNRLGIDNFEFIRVLGKGSFGKVMLARVKETGDLYAVKVLKKDVILQDDDVECTMTEKRILSLARNHPFLTQLFC
CFQTPDRLFFVMEFVNGGDLMFHIQKSRRFDEARARFYAAEIISALMFLHDKGIIYRDLKLDNVLLDHEGHCKLADFGMCKEGICNGVTTATFCGTPDYIAPEILQE
MLYGPAVDWWAMGVLLYEMLCGHAPFEAAENEDDLFEAILNDEVVYPTWLHEDATGILKSFMTKNPTMRLGSLTQGGEHAILRHPFFKEIDWAQLNHRQIEPPFRP
RIKSREDVSNFDPDFIKEEPVLTPIDEGHLPMINQDEFRNFSYVSPELQP

>H19102_H_SEQID#AA_73
MLMGFCRLEEAGLVSRSIRERNCLYNWDSRFSRERRQRLGMGAVSCRQGQHTQQGEHTRVAVPHKGGNIRGPWARGWKSLWTGLGTIRSDLEELWELRGHHYL
HQESLKPAPVLVEKPLPEWPVPQFINLFLPEFPIRPIRGQQQLKILGLVAKGSFGTVLKVLDCTQKAVFAVKVVPKVKVLQRDTVRQCKEEVSIQRQINHPFVHSLGD
SWQGKRHLFIMCSYCSTDLYSLWSAVGCFPEASIRLFAAELVLVLCYLHDLGIMHRDVKMENILDERGHLKLTDFGLSRHVPQGAQAYTICGTLQYMAPEVLSG
GPYNHAADWWSLGVLLFSLATGKFPVAAERDHVAMLASVTHSDSEIPASLNQGLSLLHELLCQNPLHRLRYLHHFQVHPFFRGVAFDPELLQKQPVNFVTETQA
TQPSSAETMPFDDFDCDLESFLLYPIPA

>MSK1_H_SEQID#AA_74
MEBBGGSSGGAAGTSADGGGDGEQLLTVKHELRTANLTGHAEKVGIENFELLKVLGTGAYGKVFLVRKISGHDTGKLYAMKVLKKATIVQKAKTTEHTRTERQV
LEHIRQSPFLVTLHYAFQIETKLHLLDYINGGELFTHLSQRERFTEHEVQIYVGEIVLALEHLHKLGIIYRDIKLENLLLDSNGHVVLTDFGLSKEFVADETERAYSFC
GTIEYMAPDIVRGGDSGHDKAVDWWSLGVLMYELLTGASPFTVDGEKNSQAEISRRILKSEPPYPQEMSALAKDLIQRLLMKDPKKRLGCGPRDADEIKEHLFFQK
INWDDLAAKKVPAPFKPVIRDELDVSNFABEFTEMDPTYSPAALPQSSEKLFQGYSFVAPSILFKRNAAVIDPLQFHMGVERPGVTNVARSAMMKDSPFYQHYDLD
LKDKPLGEGSFSICRKCVHKKSSNQAFAVKIISKRMEANTQKEITALKLCEGHPNIVKLHEVFHDQLHTFLVMELLNGGELFERIKKKKHFSETEASYIMRKLVSAVS
HMHDVGVVHRDLKPENLLFTDENDNLEIKIIDFGFARLKPPDNQPLKTPCFTLHYAAPELLNQNGYDESCDLWSLGVILYTMLSGQVPFQSHDRSLTCTSAVEIMK
KIKKGDFSFEGEAWKNVSQEAKDLIQGLLTVDPNKRLKMSGLRYNEWLQDGSQLSSSNPLMTPDILGSSGAAVHTCVKATFHAFNKYKREGFCLQNVDKAPLAKR
RKMKKTSTSTETRSSSSESSHSSSHGKTTPTKTLQPSNPADSNNPETLFQFSDSVA

>YANK3_H_SEQID#AA_75
MRSGAERRGSSAAASPGSPPPGRARPAGSDAPSALPPPAAGQPRARDSGDVRSQPRPLFQWSKWKKRMGSSMSAATARRPVFDDKEDVNFDHFQILRAIGKGSFG
KVCIVQKRDTEKMYAMKYMNKQQCIERDEVRNVFRELEILQEIEHVFLVNLWYSFQDEEDMFMVVDLLLGGDLRYHLQQNVQFSEDTVRLYICEMALALDYLR

GQHIIHRDVKPDNILLDERGHAAHLTDFNIATIIKDGERATALAGTKPYMAPEIFHSFVNGGTGYSFEVDWWSVGVMAYELLRGWRPYDIHSSNAVESLVQLFSTVS
VQYVPTWSKEMVALLRKLLTVNPEHRLSSLLQDVQAAPALAGVLWDHLSEKRVEPGFVPNKGRLHCDPTFELEEMILESRPLHKKKKRLAKNKSRDNSRDSSQSEN
DYLQDCLDAIQQDFVIFNREKLKRSQDLPREPLPAPESRDAAEPVEDEAERSALPMCGPICPSAGSG

>MARK2_H_SEQID#AA_76
MSSARTPLPTLNERDTEQPTLGHLDSKPSSKSNMIRGRNSATSADEQPHIGNYRLLKTIGKGNFAKVKLARHILTGKEVAVKIIDKTQLNSSSLQKLFREVRIMKVLN
HPNIVKLFEVIETEKTLYLVMEYASGGEVFDYLVAHGRMKEKEARAKFRQIVSAVQYCHQKFIVHRDLKAENLLLDADMNIKIADFGFSNEFTTGNKLDTFCGSPP
YAAPELFQGKKYDGPEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLLKKFLILNPSKRGTLEQIMKDRWMNVGHEDDELKPYVEP
LPDYKDPRRTELMVSMGYTREBIQDSLVGQRYNEVMATYLLLGYKSSELEGDTITLKPRPSADLTNSSAQFPSHKVQRSVSANPKQRRFSDQAGPAPTSNSYSKKT
QSNNAENKRPEBDRESGRKASSTAKVPASPLPGLERKKTTPTPSTNSVLSTSTNRSRNSPLLERASLGQASIQNGKDSLTMPGSRASTASASAAVSAARPRQHQKSM
SASVHPNKASGLPPTESNCEVPRPSTAPQRVPVASPSAHNISSSGAPDRTNFPRGVSSRSTFHAGQLRQVRDQQNLPYGVTPASPSGHSQGRRGASGSIFSKFTSKF
VRRNLSFRFARRNLNEPESKDRVETLRPHVVGSGGNDKEKEFREAKPRSLRFTWSMKTTSSMEPNEMMREIRKVLDANSCQSELHEKYMLLCMHGTPGHEDFV
QWEMEVCKLPRLSLNGVRFKRISGTSMAFKNIASKIANELKL

>NuaK2_H_SEQID#AA_77
MLCFTARSGTAVAPRPGAGPVPRASALPAQPDSPAALAHLLLAMESLVFARRSGPTPSAAELARPLAEGLIKSPKPLMKKQAVKRHHHKHNLRHRYEFLETLGKG
TYGKVKKARESSGRLVAKSRKDKIKDEQLMHIRREIEMSSLNHPHIAIHEVFENSSKIVTVMEYASRGDLYDYISERQQLSEREARHFFRQIVSAVHYCHQNRV
VHRDLKLENILLDANGNIKIADFGLSNLYHQGKFLQTFCGSPLYASPEIVNGKPYTGPEVDSWSLGVLLYLVHGTMPFDGHDHKILVKQISNGAYREPPKPSDACG
LIRWLLMVNPTRRATLEDVASHWWVNWGYATRVGEQEAPHEGGHPGDSARASMADWLRRSSRPLLENGAKVCSFFKQHAPGGGSTTPGLERQHSLKKSRKEN
DMAQSLHSDTADDTAHRPGKSNLKLPKGILKKKVSASAEGVQEDPPELSPIPASPGQAAPLLPKKGILKKPRQRESGYYSSPEPSESGELLDAGDVFVSGDPKEQKP
PQASGLLLHRKGILKLNGKFSQTALELAAPTTFGSLDELAPPRPLARASRPSGAVSEDSILSSESFDQLDLPERLPEPPLRGCVSVDNLTGLEEPPSEGPGSCLRRWRQ
DPLGDSCFSLTDCQEVTATYRQALRVCSKLT

>BRSK2_H_SEQID#AA_78
MTSTGKDGGAQHAQYVGPYRLEKTLGKGQTGLVKLGVHCVTCQKVAIKIVNREKLSESVLMKVEREIAILKLIEHPHVLKLHDVYENKKYLYLYLVLEHVSGGELFD
YLVKKGRLTPKEARKFRQIISALDFCHSHSICHRDLKPENLLLDEKNNIRIADFGMASLQVGDSLLETSCGSPHYACPEVIRGEKYDGRKADVWSCGVILFALLVG
ALPFDDDNLRQLLEKVKRGVFHMPHFIPPDCQSLLRGMIEVDAARRLTLEHIQKHIWYIGGKNEPEPEQPIPRKVQIRSLPSLEDIPPDVLDSMHSLGCFRDRNKLLQ
DLLSEENQEKMIYFLLLDRKERYPSQEDEDLPPRNEIDPPRKRVDSPMLNRHGKRRPERKSMEVLSVTDGGSPVPARRAIEMAQHGQRSRSIGASSGLSTSPLLSSP
RVTPHPSPRGSPLPTPKGTPVHTPKESPAGTPNPTPPSSPSVGGVPWRARLNSIKNSFLGSPRFHRRKLQVPTPEEMSNLTPESSPELAKKSWFGNFISLKEEQIFVVI
KDKPLSSIKADIVHAFLSIPSLSHSVISQTSFRAEYKATGGPAVFQKPVKFQVDITYTEBGEAQKENGIYSVTFTLLSGPSRRFKRVVETIQAQLLSTHDPPAAQHLSEP
PPPAPGLSWGAGLKGQKVATSYESSL

>MARK4_H_SEQID#AA_79
MSSRTVLAPGNDRNSDTHGTLGSGRSSDKGPSWSSRSLGARCRNSIASCPEEQPHVGNYRLLRTIGKGNFAKVKLARHILTGREVAIKIIDKTQLNPSSLQKLFREVR
IMKGLNHPNIVKLFEVIETEKTLYLVMEYASAGEVFDYLVSHGRMKEKEARAKFRQIVSAVHYCHQKNIVHRDLQAENLLLDAEANIKIADFGFSNEFTLGSKLDT
FCGSPPYAAPELFQGKKYDGPEVDIWSLGVILYTLVSGSLPFDGHNLKELRERVLRGKYRVPFYMSTDCESILRRFLVLNPAKRCTLEQIMKDKWINIGYEGEELKP
YTEPEEDFGDTKRIEVMVGMGYTREBIKESLTSQKYNEVTATYLLLGRKTEBGGDRGAPGLALARVRAPSDTTNGTSSSKGTSHSKGQRSSSSTYHRQRHSDFCG
PSPAPLHPKRSPTSTGEAELKEERLPGRKASCSTAGSGSRGLPPSSPMVSSAHNPNKAEIPERRKDSTSTPNNLPPSMMTRRNTYVCTERPGAERPSLLPNGKENSSGT
PRVTPASPSSHSLAPPSGERSRLARGSTIRSTFHGGQVRDRRAGGGGGGGVQNGPPASPTLAHEAAPLPAGRPRPTTNLFTKLTSKLTRRVADEPERIGGPEVTSCHL

PWDQTETAPRLLRFPWSVKLTSSRPPEALMAALRQATAAARCRCRQPQPFLLACLHGGAGGPEPLSHFEVEVCQLPRPGLRGVLFRRVAGTALAFRTLVTRISNDL
EL

>DCAMKL2_H_SEQID#AA_80
MASTRSIELEHFEERDKRPRPGSRRGAPSSSGGSSSSGPKGNGLIPSPAHSAHCSFYRTRTLQALSSEKKAKKARFYRNGDRYFKGLVFAISSDRFRSFDALLIELTRS
LSDNVNLPQGVRTIYTIDGSRKVTSLDELLEGESYVCASNEPFRKVDYTKNIPNWSVNKGGTSRALAAASSVKSEVKSKDFIKPKLVTVIRSGVKPRKAVRILLN
KKTAHSFEQVLTDITEAIKLDSGVVRLCTLDGKQVTCLQDFFGDDDVFIACGPEKFRYAQDDFVLDHSECRVLKSSYSRSSAVKYSGSKSPGPSRRSKSPASVNGT
PSSQLSTPKSTKSSSSSPTSPGSFRGLKQISAHGRSSSNVNGGPELDRCISPEGVNGNRCSESSTLLEKYKIGKVIGDGNFAVVKECIDRSTGKEFALKIIDKAKCCGKE
HLIENEVSILRRVKHPNIIMLVBEMETATEHLRLVMELVKGGDLFDAITSSTKYTERDGSAMVYNLANALRYLHGLSIVHRDIKPENLLVCEYPDGTKSLKLGDFGLA
TVVEGPLYTVCGTPTYVAPEIIAETGYGLKVDIWAAGVITYILLCGFPPFRSENNLQEDLFDQILAGKLEFPAPYWDNITDSAKELISQMLQVNVEARCTAGQILSHP
WVSDDASQENNMQAEVTGKLKQHFNNALPKQNSTTTGVSVIMNTALDKEGQIFCSKHCQDSGRPGMEPISPVPPSVEEIPVPGEAVPATTPPESPTPHCPPAAPGGE
RAGTWRRHRD

>PIM2_H_SEQID#AA_81
LAPIGCGTAGAASGLIKARWAGLSPRVTLPGGACPGSVALGLARAANLNAAPSAGASGPPVSLLSTLAPPSPGSPAALPRASTPCGLSGFSGLNLRSATSMLTKPLQ
GPPAPPGTPTPPPGGKDREAFEAEYRLGPLLGKGGFGTVFAGHRLTDRLQVAIKVIPRNRVLGWSPLSDSVTCPLEVALLWKVGAGGHPGVIRLLDWFETQEGFM
LVLERPLPAQDLFDYITEKGPLGEGPSRCFGQVVAAIQHCHSRGVVHRDIKDENLIDILRRGCAKLIDFGSGALLHDEPYTDFDGTRVYSPPEWISRHQYHALPATV
WSLGILLYDMVCGDIPFERDQEILEAELHFPAHVSPDCCALIRRCLAPKPSSRPSLEIILLDPWMQTPAEDVTPQPLQRRPCPFGLVLATLSLAWPGLAPNGQKSHPM
AMSQG

>PIM3_H_SEQID#AA_82
MLLSKFGSLAHLCGPGGVDHLPVKILQPAKADKESFEKAYQVGAVLGSGGFGTVYAGSRIADGLPVAVKHVVKERVTEWGSLGGATVPLEVVLLRKVGAAGGA
RGVIRLLDWFERPDGFLLVLERPEPAQDLFDFITERGALDEPLARRFFAQVLAAVRHCHSCGVVHRDIKDENLLVDLRSGELKLIDFGSGALLKDTVYTDFDGTRVY
SPPEWIRYHRYHGRSATVWSLGVLLYTLVVAHLPFDDTNLKKLLRETQKEVTFPANHTISQECKNLILQMLRQATKRATILDHKDSWVLKFQPEQPTHEIRLLEAMCQLHNTIK
ASTTSSSESL

>TSSK4_H_SEQID#AA_83
MGKGDVLEAAPTTTAYHSLMDEYGYEVGKAIGHGSYGSVYEAFYTKQKVMVAVKIISKKKASDDYLNKFLPREIQVMKVLRHKYLINFYRAEBSTSRVYIILELAQ
GGDVLEWIQRYGACSEPLAGKWFSQLTLGIAYLHSKSIVHRDLKLENLLLDKWENVKISDFGFAKMVPSNQPVGCSPSYRQVNCFSHLSQTYCGSFAYACPEILRG
LPYNPFLSDTWSMGVILYTLVVAHLPFDDTNLKKLLRETQKEVTFPANHTISQECKNLILQMLRQATKRATILDHKDSWVLKFQPEQPTHEIRLLEAMCQLHNTIK
QHQSLQITT

>CKI2_H_SEQID#AA_84
MSGGGEQLDILSVGILVKERWKVLRKIGGGGFGEIYDALDMLTRENVALKVESAQQPKQVLKMEVAVLKKLQGKDHVCRFIGCGRNDRFNYVVMQLQGRNLAD
LRRSQSRGTFTISTTLRLGRQILESIESIHSVGFLHRDIKPSNFAMGRFPSTCRKCYMLDFGLARQFTNSCGDVRPPRAVAGFRGTVRYASINAHTNREMGRHDDLWS
LFYMLVEFVVGQLPWRKIKDKEQVGSIKERYDHRLMLKHLPPERSIFLDHISSLDYFTKPDYQLLTSVFDNSIKTFGVIESDPFDWEKTGNDGSLTTTTSTTPQLHT
RLTPAAIGIANATPIPGDLLRENTDEVFPDEQLSDGENGIPVGVSPDKLPGSLGHPRPQEKDVWEEMDANKNKIKLGICKAATEEENSHGQANGLINAPSLGSPIRVR
SEITQPDRDIPLVRKLRSIHSFELEKRLTLEPKPDTDKFLETCLEKMQKDTSAGKESILPALLHKPCVPAVSRTDHIWHYDEEYLPDASKPASANTPEQADGGGSNGFI
AVNLSSCKQEIDSKEWVTVDKEQDLQDFRTNEAVGHKTTGSPSDEEPEVLQVLEASPQDEKLQLGPWAENDHLKKETSGVVLALSAEGPPTAASEQYTDRLELQP

GAASQFIAATPTSLMBAQAEGPLTAITIPRPSVASTQSTSGSFHCGQQPEKKDLQPMEPTVELYSPRENFSGLVVTEGEPPSGGSRTDLGLQIDHIGHDMLPNIRESNK
SQDLGPKELPDHNRLVVRFFENLPGETEEKSILLESDNEDEKLSRGQHCIEISSLPGDLVIVEKDHSAITEPLDVTKTQTFSVVPNQDKNNEMKLLTVGTSEISSRDID
PHVEGQIGQVAEMQKNKISKDDDIMSEDLPGHQGDLSTFLHQEGKREKITPRNGELFHCVSENEHGAPTRKDMVRSSFVTRHSRIPVLAQEIDSTLESSSPVSAKEKL
LQKKAYQPDLVKLLVEKRQFKSFLGDLSSASDKLLEEKLATVPAPFCEEEVLTPFSRLTVDSHLSRSAEDSFLSPIISQSRKSKIPRPVSWVNTDQVNSSTSSQFFRPP
PGKPPTRPGVEARLRRYKVLGSSNSDSDLFSRLAQILQNGSQKPRSTTQCKSPGSPHNPKTPPKSPVVPRRSPSASPRSSSLPRTSSSSPSRAGRPHHDQRSSSPHLGRS
KSPPSHSGSSSSRRSCQQEHCKPSKNGLKGSGSLHHHSASTKTPQGKSKPASKLSR

>PCTAIRE3_H_SEQID#AA_85
MIMNKMKFNKRRFSLSVPRTETIEESLAEFTEQFNQLHNRRNENLQLGPLGRDPPQECSTFSPTDSGEEPGQLSPGVQFQRRQNQRRFSMEVRASGALPRQVAGCTH
KGVHRRAAALQPDFDVSKRLSLPMDIRLPQEFLQKLQMESPDLPKPLSRMSRRASLSDIGFGKLETYVKLDKLGEGTYATVFKGRSKLTENLVALKEIRLEHEEGAP
CTAIREVSLLKNLKHANIVTLHDILIHTDRSLTLVFEYLDSDLKQYLDHCGNLMSMHNVKIFMFQLLRGLAYCHHRKILHRDLKPQNLLINERGELKLADFGLARAK
SVPTKTYSNEVVTLWYRPPDVLLGSTEYSTPIDMWGVGCIHYEMATGRPLFPGSTVKEELHLIFRLLGTPTEETWPGVTAFSEFRTYSFPCYLPQPLINHAPRLDTDG
IHLLSSLLLYESKSRMSAEAALSHSYFRSLGERVHQLEDTASIFSLKEIQLQKDPGYRGLAFQQPGRGKNRRQSIF

>PFTAIRE2_H_SEQID#AA_86
MGQELCAKTVQPGCSCYHCSEGGEAHSCRRSQPETTEAAFKLTDLKEASCSMTSFHPRGLQAARAQKFKSKRPRSNSDCFQEEDLRQGFQWRKSLPFGAASSYLN
LEKLGEGSYATVYKGISRINGQLVALKVISMNABEGVPFTAIREASLLKGLKHANIVLLHDIIHTKETLTFVFEYMHTDLAQYMSQHPGGLHPHNVRLFMFQLLRGL
AYHHQHVLHRDLKPQNLLISHLGELKLADFGLARAKSIPSQTYSSEVVTLWYRPPDALLGATEYSSEIDWGAGCIFIEMFQGQPLFPGVSNILEQLEKIWEVLGVP
TEDTWPGVSKLPNYNPEWFPLTPRSLHVVWNRLGRVPEAEDLASQMLKGFPRDRVSAQEALVHDYFSALPSQLYQLPDEESLFTVSGVRLKPEMCDLLASYQKG
HHPAQFSKCW

>ERK7_H_SEQID#AA_87
MCTVVDPRIVRRYLLRRQLGQGAYGIVWKAVDRRTGEVVAIKKIFDAFRDKTDAQDMGFLLAPPTHTPVFLSLQRTFREITLLQEFGDHPNIISLLDVIRAENDRDIY
LVFEFMDTDLNAVIRKGGLLQDVHVRSIFYQLLRATRFLHSGHVVHRDQKPSNVLLDANCTVKLCDFGLARSLGDLPEGPEDQAVTEYVATRWYRAPEVLLSSHR
YTLGVDMWSLGCILGEMLRGRPLFPGTSTLHQLELILETPPPSEEDLLALGSGCRASVLHQLGSRPRQTLDALLPPDTSPEALDLLRRLLVFAPDKRLSATQALQHP
YVQRFHCPSDEWAREADVRPRAHEGVQLSVPEYRSRVYQMILECGSSGTSREKGPEGVSPSQAHLHKPRADPQLPSRTPVQGPRPRPQSSPGHDPAEHESPRAAK
NVPRQNSAPLLQTALLGNGERPPGAKEAPPLTLSLVKPSGRGAAPSLTSQAAAQVANQALIRGDWNRGGGVRVASVQQVPPRLPEARPGRRMFSTSALQGAQGG
ARALLGGYSQAYGTVCHSALGHLPLLEGHHV

>CKIIa-rs_H_SEQID#AA_88
MSGPVPSRARVYTDVNTHRPREYWDYESHVVEWGNQDDYQLVRKLGRGKYSEVFEAINTNNEKVVVKILKPVKKKKIKREIKILENLRGGPNIITLADIVKDPVS
RTPALVFEHVNNTDFKQLYQTFIDYDIRFYMYEILKALDYCHSMGMHRDVKPHNVMIDHEHRKLRLIDWGLAEFYHPGQEYNVRVASRYFKGPELLVDYQMYD
YSLDMWSLGCMLASMIFRKEPFFHGRDNYDQLVRIAKVLGTHDLYGYIDKYNIELDPRFNDILGRHSRKRWBRFVHRENQHLVSPEALDFLDKLLRYDHQSRLTA
REAMEHPYFYTVVKDQARMGSSSMPGGSTPVSSANVMSGISSVPTPSPLGAGSPVIAAANPLGMPVPAAAGAQQ

>DYRK4_H_SEQID#AA_89
MNTPLQSAHTHLTLVEFIAYLVSDSPFPAASTLSLPTCNFDPCPTKLGAGTMPPAAAGGQRRPHAAPPAAYPHRNKVRAAEARTSRAVRREYEAGDEASGRRGRG
GRMKSVQNARADRGMWVRERRAVVRGQENTVVFAAAASPESDFLQIVLRDIAREMVPEYKSDLQPRTQMDAKKPRKCDLTPFLVLKARKKQKFTSAKHLWV
QRKGKGVEDHADWQGGWGGQIWVGVEGKSAEGQLNLTGAFNRSPENKVSPQGPALNALEIDSSYLDPQALALPQGWQGLLFPPLLGTELVGSKLSVQIQKPPSNIKNS

RMTQVFHKNTSVTSLPFVDTKGKKNTVSFPHISKKVLLKSSLLYQVSADGLDPEKAGERPGFSPKPVLERPRIVGKSTVAAEEENQAHNQMPASELKASEIPFHPSIK
TQDPKAEEKSPKKQKVTLTAAEALKLFKNQLSPYEQSEILGYAELWFLGLEAKKLDTAPEKFSKTSFDDEHGFYLKVLHDHIAYRYEVLETIGKGSFGQVAKCLDH
KNNELVALKIIRNKKRFHQQALMELKILEALRKKDKDNTYNVHMKDFFYFRNHFCITFELLGINLYELMKNNNFQGFSLSIVRRFTLSVLKCLQMLSVEKIIHCDL
KPENIVLYQKGQASVKVDDFGSSCYEHQKVYTYIQSRFYRSPEVILGHPYDVAIDMWSLGCITAELYTGYPLFPGENEVEQLACIMEVLGLPPAGFIQTASRRQTFFD
SKGFPKNITNNRGKKRYPDSKDLTMVLKTYDTSFLDFLRRCLVWEPSLRMTPDQALKHAWIHQSRNLKPQPRPQTLRKSNSFFPSETRKDKVQGCHHSSRKDEITK
ETTEKTKDSPTKHVQHSGDQQDCLQHGADTVQLPQLVDAPKKSEAAVGAEVSMTSPGQSKNFSLKNTNVLPPIV

>HIPK1_H_SEQID#AA_90
MASQLQVFSPPSVSSSAFCSAKKLKIEPSGWDVSGQSNDKYYTHSKTLPATQGQANSSHQVANFNIPAYDQGLLLPAPAVEHIVVTAADSSGSAATSTFQSSQTLT
HRSNVSLLEPYQKCGLKRKSEEVDSNGSVQIIEEHPPLMLQNRTVVGAAATTTVTKSSSSGEGDYQLVQHEILCSMTNSYEVLEFLGRGTFGQVAKCWKRSTK
EIVAIKILKNHPSYARQGQIEVSILSRLSSENADEYNLVRSYECFQHKNHTCLVFEMLEQNLYDFLKQNKFSPLPLKYIRPILQQVATALMKLKSLGIHADLKPENIM
LVDPVRQPYRVKVDDFGSASHVSKAVCSTYLQSRYYRAPEIILGLPFCEAIDMWSLGCVIAELFLGWPLYPGASEYDQIRYISQTQGLPAEYLLSAGTKTTRFFNRDP
NLGYPLWRLKTPEEHELETGIKSKEARKYTFNCLDDMAQVNMSTDLEGTDMLAEKADRREYTDLLKKMLTIDADKRITPLKTLNHQFVTMTHLLDFPHSNHVKSC
FQNMEICKRRVHMYDTVSQIKSPFTTHVAPNTSTNLTMSFSNQLNTVHNQASVLASSSTAAAATLSLANSDVSLLNYQSALYPSSAAPVPGVAQQGVSLQPGTTQI
CTQTDPFQQTFIVCPPAFQTGLQATTKHSGFPVRMDNAVPIVPQAPAAQPLQIQSGVLTQGSCTPLMVATLHPQVAITTPQYAVPFTLSCAAGRPALVEQTAAVLQA
WPGGTQQILLPSTWQQLPGVALHNSVQPTAMIPEAMGSGQQLADWRNAHSHGNQYSTIMQQPSLLTNHVTLATAQPLNVGVAHVVRQQQSSLPSKKNKQSAPV
SSKSSLDVLPSQVYSLVGSSPLRTTSSYNSLVPVQDHQPIIPDTPSPPVSVITIRSDTDEEEDNKYKPSSSGLKPRSNVISYVTVNDSPDSDSLSSPYSTDTLSALRGN
SGSVLEGPGRVVADGTGTRITIVPPLKTQLGDCTVATQASGLLSNKTKPVASVSGQSSGCCITPTGYRAQRGGTSAAQPLNLSQNQQSSAAPTSQERSSNPAPRRQQ
AFVAPLSQAPYTFQHGSPLHSTGHPHLAPAPAHLPSQAHLYTYAAPTSAAALGSTSSIAHLFSPQGSSRHAAAYTTHPSTLVHQVPVSVGPSLLTSASVAPAQYQHQ
FATQSYIGSSRGSTIYTGYPLSPTKISQYSYL

>HIPK4_H_SEQID#AA_91
MSTIQSETDCYDIIEVLGKGTFGEVAKGWRRSTGEMVAIKILKNDAYRNRIIKNELKLLHCMRGLDPEEAHVIRFLEFFHDALKFYLVFELLEQNLFEFQKENNFAPL
PARHIRTVTLQVLTALARLKELAIIHADLKPENIMLVDQTRCPFRVKVDIFGSASIFSEVRYVKEPYIQSRFYRAPEILLGLPFCEKVDVWSLGCVMAELHLGWPLYP
GNNEYDQVRYICETQGLPKPHLIHAACKAHHFFKRNPHPDAANPWQLKSSADYLAETKVRPLERRKYMLKSLDQETVNGGSVASRLTFPDREALAEHADLKSM
VELIKRMLTWESHERISPSAALRHPFVSMQQLRSAHETTHYYQLSLRSYRLSLQVEGKPPTPVVAAEDGTPYYCLAEEKEAAGMGSVAGSSPFFREEKAPGMQRAI
DQLDDLSLQEAGHGLWGETCTNAVSDMMVPLKAAITGHHVPDSGPEPILAFYSSRLAGRHKARKPPAGSKSDSNFSNLIRLSQVSPEDDRPCRGSSWEEGEHLGAS
AEPLAILQRJDEDGPNIDNMTMEAAERPDPELFDPSSCPGEWLSEPDCTLESVRGPRAQGLPPRRSHQHGPPRGATSFLQHVTGHH

>BIKE_H_SEQID#AA_92
MKKFSRMPKSEGGSGGGAAGGGAGGGAGGAGAGAGCGSGGGSSVGVRVFAVGRHQVTLEESLAEGGFSTVFLVRTHGGIRCALKRMYVNNMPDLNVCKREITIMKELS
GHKNIVGYLDCAVNSISDNVWEVLILMEYCRAGQVVNQMNKKLQTGFTEPEVLQIFCDTCEAVARLHQCKTPIIHRDLKVENILLNDGGNYVLCDFGSATNKFLN
PQKDGVNVVEEEIKKYTTLSYRAPEMINLYGGKPITTKADIWALGCLLYKLCFFTLPFGESQVAICDGNFTIPDNSRYSRNIHCLIRFMLEPDPEHRPDIFQVSYFAFK
FAKKDCPVSNINNSSIPSALPEPMTASEAAARKSQIKARITDTIGPTETSIAPRQRPKANSATTATPSVLTIQSSATPVKVLAPGEFGNHRPKGALRPGNGPEILLGQGP
PQQPPQQHRVLQQLQQGDWRLQQLHLQHRHPHQQQGGQQQQQQQQQQQQHHHHHHHHLLQDAYMQQYQHATQQQMLQQQFLMHSVYQ
PQPSASQYPTMMPQYQQAFFQQQMLAQHQPSQQQASPEYLTSPQEFSPALVSYTSSLPAQVGTIMDSSYSANRSVADKEAIANFTNQKNISNPPDMSGWNPFGEDN
FSKLTEEBLLDREFDLLRSNRLEERASSDKNVDSLSAPHNHPPEDPFGSVPFISHSGSPEKKAEHSSINQENGTANPIKNGKTSPASKDQRTGKKTSVQGQVQKGNDE
SESDFFESDPPSPKSSEEBEQDDEEVLQGEQGDFNDDDTEPENLGHRPLLMDSEDEEEEKHSSDSDYEQAKAKYSDMSSVYRDRSGSGPTQDLNTILLTSAQLSSDV
AVETPKQEFDVFGAVPFFAVRAQQPQQEKNEKNLPQHRFPAAGLEQEEFDVFTKAPFSKKVNVQECHAVGPEAHTIPGYPKSVDVFGSTPFQPFLTSTSKSESNEDL

FGLVPFDEITGSQQQKVKQRSLQKLSSRQRRTKQDMSKSNGKRHHGTPTSTKKTLKPTYRTPERARRHKKVGRRDSQSSNEFLTISDSKENISVALTDGKDRGNVL
QPEESLLDPFGAKPFHSPDLSWHPPHQGLSDIRADHNTVLPGRPRQNSLHGSFHSADVLKMDDFGAVPFTELVVQSITPHQSQSQPVELDPFGAAPFPSKQ

>NEK10_H_SEQID#AA_93
MPDQDKKVKTTEKSTDKQQEITRDYSDLKRLRCLLNVQSSKQQLPAINFDSAQNSMTKSEPAIRAGGHRARGQWHESTEAVELENFSINYKNERNFSKHPQRKLF
QEIFTALVKNRLISREWVNRAPSIHFLRVLICLRLLMRDPCYQEILHSLGGIENLAQYMEIVANEYLGYGEEQHTVDKLVNMIYIFQKLAAVKDQREWVTTSGAHK
TLVNLLGARDTNVLLGSLLALASLAESQECREKISELNIVENLLMILHEYDLLSKRLTAELLRLLCAEPQVKEQVKLYEGIPVLLSLLHSDHLKLLWSIVWILVQVCE
DPETSVEIRIWGGIKQLLHILQGDRNFVSDHSSIGSLSSANAAGRIQQLHLSEDLSPREIQENTFSLQAACCAALTELVLNDTNAHQVVQENGVYTIAKLILPNKQKN
AAKSNLLQCYAFRALRFLFSMERNRPLFKRLFPTDLFEIFDIGHYVRDISAYEELVSKLNLLVEDELKQIAENIESINQNKAPLKYIGNYAILDHLGSGAFGCYYKVR
KHSGQNLLAMKEVNLHNPAFGKDKKDRDSSVRNIVSELTIIKEQLYHPNIVRYYKTFLENDRLYIVMELIEGAPLGEHFSSLKEKHHFTEERLWKIFIQLCLALRYL
HKEKRIVHRDLTPNNIMLGDKDKVTVTDFGLAKQKQENSKLTSVVGTILYSCPEVLKSEPYGEKADVWAVGCILYQMATLSPPFYSTNMLSLATKIVEAVYBPVPE
GIYSEKVTDTISRCLTPDAEARPDIVEVSSMISDVMMKYLDNLSTSQLSLEKKLERERRRTQRYFMEANRNTVTCHHELAVLSHETFEKASLSSSSSGAASLKSELSE
SADLPPEGFQASYGKDEDRACDEILSDDNFNLENAEKDTYSEVDDELDISDNSSSSSSSPLKESTFNILKRSFSASGGERQSQTRDFTGGTGSRPRPGPQMGTFLWQA
SAGIAVSQRKVRQISDPIQQILQLHKIIYTQLPPALHHNLKRRVIERFKKSLFSQQSNPCNLKSEIKKLSQGSPEPIEPNFFTADYHLLHRSSGGNSLSPNDPTGLPTSIE
LEEGITYEQMQTVIEEVLEESGYYNFTSNRYHSYPWGTKNHPTKR

>pNEK5_H_SEQID#AA_94
MQIVGSPGPGAAWPVKRVVFPNGEQFLLSVATKKVICLCLGKAGRKVLAKKLSPLETMDKYDVIKAIGQGAFGKAYLAKGKSDSKHCVIKENFEKMPIQEKEAS
KKEVILLEKMKHPNIVAFFNSFQENGRLFIVMEYCDGGDLMKRINRQRGVLFSEDQILGWFVQISLGLKHIHDRKILHRDIKAQNFLSKNGMVAKLGDFGIARVLN
NSMELARTCIGTPYYLSPBICQNKPYNNKTDIWSLGCVLYELCTLKHPFEGNNLQQLVLKICQAHFAPISPGFSRELHSLISQLFQVSPRDRPSINSILKRPFLENLIPKY
LTPEVIQEBFSHMLICRAGAPASRHAGKVVQKCKIQKVRFQGKCPPRSRISVPIKRNAILHRNEWRPPAGAQKARSIKMERPKIAAVCGHYDYYAQLDMLRRRA
HKPSYHPPQENTGVEDYGQETRHGPSPSQWPAEYLQRKFEAQQYKLKVEKQLGLRPSSAEPNYNQRQELRSNGEEPRFQELPFRKNEMKEQEYWKQLEIRQQY
HNDMKEIRKKMGREPEENSKISHKTYLVKKSNLPVHQDASEGEAPVQDIEKDLKQMRLQNTKESKNPEQKYKAKKGVKFEINLDKCISDENILQEEAMDIPNETL
TFEDGMKFKEYECVKEHGDYIDKAFEKLHCPEAGFSTQTVAAVGNRRQWDGGAPQTLLQMMAVADITSTCPTGPDNGQVIVEGIPGNRKQWRHEAPGTLMSVL
AAAHLTSSSFSADEEFAMGTLKQWLPKEEDEGKVEMVSGIEVDBEQLEPRSDDDDTNFEESEDELRDEVVEYLEKLATFKGEEKTEEASSTSKDSRKSREREGISM
QKSEELREGLENISTTSNDHICITDEDQGTSTTSQNIQV

>NEK1_H_SEQID#AA_95
MEKYVRLQKIGBGSFGKAILVKSTEDGRQYVIKEINISRMSSKEREESRREVAVLANMKHPNIVQYRESFEENGSLYTVMDYCEGGDLFKRINAQKGVLFQEDQILD
WFVQICLALKHVHDRKILHRDIKSQNFLTKDGTVQLGDFGIARVLNSTVELARTCIGTPYYLSPEICENKPYNNKSDIWALGCVLYELCTLKHAFEAGSMKNLVLK
IISGSFPPVSLHYSYDLRSLVSQLFKRNPRDRPSVNSILEKGFIAKRIEKFLSPQLIABEFCLKTFSKFGSQPTPAKRPASGQNSISVMPAQKITKPAAKYGIPLAYKKYG
DKKLHEKKPLQKHKQAHQTPEKRVNTGBERRKISEEAARKRRLEFIEKEKKQKDQIISLMKAEQMKRQEKERLERINRAREQGWRNVLSAGGSGEVKAPFLGSGG
TIAPSSFSSRGQYBHYHAIFDQMQQQRAEDNEAKWKREIYGRGLPERGILPGVRPGFPYGAAGHHHFPDADDIRKTLKRLKAVSKQANANRQKGQLAVERAKQVE
EFLQRKREAMQNKARABGHMGILQNLAAMYGGRPSSSRGGKPRNKEEVYLARLRQIRLQNFNERQQIKAKLRGEKKEANHSEGQEGSEEADMRKKIESLKAH
ANARAAVLKEQLERKRKEAYEREKKVWEEHLVAKGVKSSDVSPPLGQHETGGSPSKQQMRSVISVTSALKEVGVDSSLTDTRETSEEMQKTTNAISSKREILRRL
NENLKAQEDEKGKQNLSDTFEINVHEDAKEHEKEKSVSSDRKKWEAGGQLVIPLDELTLDTSFSTTERHTVGEVIKLGPNGSPRRAWGSPTDSVLKILGEAELQL
QTELLENTTIRSEISPEGEKYKPLITGEKKVQCISHBINPSAIVDSPVETKSPEFSEASPQMSLKLEGNLEEPDDLETEILQEPSGTNKDESLPCTITDVWISEEKETKETQ
SADRITIQENEVSEDGVSSTVDQLSDIHIBPGTNDSQHSKCDVDKSVQPEPFHKVVHSEHLNLVPQVQSVQCSPEESFAFRSHSHLPPKNKNKNSLLIGLSTGLFDAN
NPKMLRTCSLPDLSKLFRTLMDVPTVGDVRQDNLEIEDIEIEDENIKEGPSDSEDIVFEETDTDLQELQASMEQLLREQPGEEYSEEEESVLKNSDVEPTANGTDVADE

DDNPSSESALNEEWHSDNSDGEIASECECDSVFNHLEELRLHLEQEMGFEKFFEVYEKIKAIHEDEDENIEICSKIVQNILGNEHQHLYAKILHLVMADGAYQEDND
E

>NEK3_H_SEQID#AA_96
MDDYMVLRMIGEGSFGRALLVQHESSNQMFAMKEIRLPKSFSNTQNSRKEAVLLAKMKHPNIVAFKESFEAEGHLYIVMEYCDGGDLMQKIKQQKGKLFPEDMI
LNWFTQMCLGVNHIHKKRVLHRDIKSKNFLTQNGKVKLGDFGSARLLSNPMAFACTYVGTPYYVPPEIWENLPYNNKSDIWSLGCILYELCTLKHPFQANSWKN
LILKVCQGCISPLPSHYSYELQRLVKQMFKRNPSHRPSATTLLSRGIVARLVQKCLPEIMEYGEEVLEEIKNSKHNTPRKKTNPSRIRIALGNEASTVQEEEQDRKGS
HTDLESINENLVESALRRVNREEKGNKSVHLRKASSPNLHRRQWEKNVPNTALTALENASILTSSLTAEDDRGGSVIKYSKNTTRKQWLKETPDTLLNLKNADLSL
AFQTYTIYRPGSEGFLKGPLSEETEASDSVDGGHDSVILDPERLEPGLDEEDTDFEEEDDNPDWVSELKKRAGWQGLCDR

>SGK069_H_SEQID#AA_97
MPGKQSEEGPAEAGASEDSEEEGLGGLTLEELQQGQEAARALEDMMTLSAQTLVRAEVDELYEEVRPLGQGRYGRVLLVTHRQKGTPLALKQLPKPRTSLRGFL
YBFCVGLSLGAHSAIVTAYGIGESAHSYSFLTEPVLHGDLMAFIQPKVGLPQPAVHRCAAQLASALEYTHARGLVYRDLKPENVLVCDPACRRFKLTDFGHTRPRG
TLLRLAGPPIPYTAPELCAPPPLPEGLPIQPALDAWALGVLLFCLLTGYFPWDRPLAEADPFYEDFLIWQASGQPRDRPQPWFGLAAAADALLRGLLDPHPRRRSAVI
AIREHLGRPWRQREGEABAAVGAVEEEAGQ

>SGK110_H_SEQID#AA_98
MDSETHSGMHRGRGRWRERPGWAGGLCGLRMHPHSGLGAPGLLPQTGAGGASVAVTPNLSRTQKQVARVREDTATALQRLVELTTSRVTPVRSLRDQYHLIRK
LGSGSYGRVLLAQPHQGGPAVALKLLRRDLVLRSTFLREFCVGRCVSAHPGLLQTLAGPLQTPRYFAFAQEYAPCGDLSGMLQERGLPELLVKRVVAQLAGALDF
LHSRGLVHADVKPDNVLVFDPVCSRVALGDLGLTRPEGSPTAPPVPLPTAPPELCLLPPDTLPRPAVDSWGLGVLLFCAATACFPWDVALAPNPEFEAFAGWV
TTKPQPPQPPPWDQFAPPALALLQGLLDLPETRSPPLAVLDFLGDDWGLQGNREGPGVLGSAVSYEDREEGGSSLEEWTDEGDDSKSGGRTGTDGGAP

>NRBP2_H_SEQID#AA_99
MAAPEPAPRRARERERDESEDEDESDILEESPCGRWQKRREQVNQGNMPGLQSTFLAMDTEBGVEVVVNELHFGDRKAFAAHEEKIQTVFEQLVLVDHPNIVKL
HKYWLDTSEACARVFITEYVSSGSLKQFLKKTKNHKAMNARAWKRWCTQILSALSFLHACSPPIIHGNLTSDTIFIQHNGLIKIGSVWHRIFSNALRPPTALPDDL
RSPRIAEREELRNLHFFPPEYGEVADGTAVDIFSFGMCALEMAVLEIQTNGDTRVTEEAIARARHSLSDPNMREFILCCLARDPARRPSAHSLLFHRVLFEVHSLKLL
AAHCFIQHQYLMPENVVEEKTKAMDLHAVLAELPRPRRPPLQWRYSEVSFMELDKFLEDVRNGIYPLMNFAATRPLGLPRVLAPPPEEVQKAKITPTPEPFDSETRK
VIQMQCNLERSEDKARWHLTLLLVLEDRLHRQLTYDLLPTDSAQDLASELVHYGFLHEDDRMKLAAFLESTFLKYRGTQA

>CNK_H_SEQID#AA_100
MBPAAGFLSPRPFQRAAAAPAPPAGPGPPSALRGPELEMLAGLPTSDPGRLITDPRSGRITYLKGRLLGKGGFARCYEATDTETGSAVAVKVTPQSRVAKPHQREKI
LNEIELHRDLQHRHIVRFSHHFEDADNIYIFLELCSRKSLAHIWKARHTLLBPEVRYYLRQILSGLKYLHQRGILHRDLKLGNFFITENMELKVGDFGLAARLEPPEQ
RKKTICGTPNYVAPEVLLRQGHGPEADVWSLGCVMYTLLCGSPPFETADLKETYRCIKQVHYTLPASLSLPARQLLAAILRASPRDRPSIDQILRHDFFTKGYTPDRL
PISSCVTVPDLTPPNPARSLFAKVTKSLFGRKKKSKNHAQERDEVSGLVSGLMRTSVGHQDARPEAPAASGPAPVSLVETAPEDSSPRGTLASSGDFEEGLTVATV
VESALCALRNCLAFMPPAEQNPAPLAQPEPLVWVSKWVDYSNKFGFGYQLSSRRVAVLFNDGTHMALSANRKTVHYNPTSTKHFSFSVGAVPRALQPQLGILRYF
ASYMEQHLMKGGDLPSVEEVEVPAPPLLLQWVKITDQALLMLFSDGTVQVNFYGDHTKLILSGWEPLLVTFVARNRSACTYLASHLRQLGCSPDLRQRLRYALRL
LRDRSPA

>SCYL2_H_SEQID#AA_101

MESMLNKLKSTVTKVTADVTSAVMGNPVTREFDVGRHIASGGNGLAWKIFNGTKKSTKQEVAVFVFDKLIDKYQKFEKDQIIDSLKRGVQQLTRLRHPRLLTVQ
HPLEESRDCLAFCTEPVFASLANVLGNWENLPSPISPDIKDYKLYDVETKYGLLQVSEGLSFLHSSVKMVHGNITPENILNKSGAWKIMGFDFCVSSTNPSEQEPKF
PCKEWDPNLPSLCLPNPEYLAPEYILSVSCETASDMYSLGTVMYAVFNKGKPIFEVNKQDIYKSFSRQLDQLSRLGSSSLTNIPEEVREHVKLLLNVTPTVRPDADQ
MTKIPFFDDVGAVTLQYFDTLFQRDNLQKSQFFKGLPKVLPKLPKRVTVQRILPCLTSEFVNPDMVPFVLPNVLLIAEECTKEEYVKLILPELGPVFKQQEPIQASNMI
LLIFLQKMDLLLTKTPPDEIKNSVLPMVYRALEAPSIQIQELCLNIIPTFANLIDYPSMKRNALIPRIKNACLQTSSLAVRVNSLVCLGKHLEYLDKWFVLDDILPFLQQIP
SKEPAVLMGILGIYKCTFTHKKLGITKEQLAGKVLPHLIPLSIENNLNLNQFNSFISVIKEMLNRLESEHKTKLEQLHIMQEQQKSLDIGNQMNVSEEMKVTNIGNQQ
IDKVFNNIGADLLTGSESENKEDGLQNKHKRASLTLEEKQKLAKEQEAQKLKSQQPLKPQVHTPVATVKQTKDLTDTLMDNMSSLTSLSVSTPKSSASSTFTSVP
SMGIGMMFSTPTDNTKRNLTNGLNANMGFQTSGFNMPVNTNQNFYSSPSTVGVTKMTLGTPPTLPNFNALSVPPAGAKQTQQRPTDMSALNNLFGPQKPKVSMN
QLSQQKPNQWLNQFVPPQGSPTMGSSVMGTQMNVIGQSAFGMQGNPFFNPQNFAQPPTTMTNSSSASNDLKDLFG

>SRPK2_H_SEQID#AA_102
MSVNSEKSSSSERPEPQQKAPLVPPPPPPPPPPPPPPPLPDPTPEPEBEILGSDDEEQEDPADYCKGGYHPVKIGDLFNGRYHVIRKLGWGHFSTVWLCWDMQGKRFV
AMKVVKSAQHYTETALDEIKLLKCVRESDPSDPNKDMVVQLIDDFKISGMNGIHVCMVFEVLGHHLLKWIIKSNYQGLPVRCVKSIIRQVLQGLDYLHSKCKIIHT
DIKPENILMCVDDAYVRRMAAEATEWQKAGAPPPSGSAVSTAPQQKPIGKISKNKKKKLKKKQKRQAELLEKRLQEIEELEREAERKIIEENITSAAPSNDQDGEYC
PEVKLKTTGLEEAAEAETAKDNGEAEDQEEKEDAEKENIEKDEDDVDQELANIDPTWIESPKTNGHIENGPFSLEQQLDEDDDEEDCPNPEEYNLDEPNAESDYT
YSSSYEQFNGELPNGRHKIPESQFPEFSTSLFSGSLEPVACGSVLSEGSPLTEQESSPSHDRSRTVSASSTGDLPKAKTRAADLLVNPLDPRNADKIRVKIADLGNAC
WVHKHFTEDIQTRQYRSEVLIGAGYSTPADIWSTACMAFELATGDYLFEPHSGEDYSRDEDHLAHIIELLGSIPRHFALSGKYSREFFNRRGELRHITKLKPWSLFDV
LVEKYGWPHEDAAQFTDFLIPMLEMVPEKRASAGECLRHPWLNS

>TLK1_H_SEQID#AA_103
MSVQSSSGSLEGPPSWSQLSTSPTPGSAAAARSLLNHTPPSGRPREGAMDELHSLDPRRQELLEARFTGVASGSTGSTGSCSVGAKASTNNESSNHSFGSLGSLSDKE
SETPEKKQSESSRGRKRKAENQNESSQGFPNLPVFQSLAYWEMGRTAGGKSIGGRGHKISDYFEYQGGNGSSPVRGIPPAIRSPQNSHSHSTPSSSVRPNSPSPTALAF
GDHPIVQPKQLSFKIIQTDLTMLKLAALESNKIQDLEKKEGRIDDLRANCDLRRQIDEQQKLLEKYKERLNKCISMSKKLLIEKSTQEKLSSREKSMQDRLRLGHFT
TVRHGASFTEQWTDGFAFQNLVKQQEWVNQQREDIERQRKLLAKRKPPTANNSQAPSTNSEPKQRKNKAVNGAENDPFVRPNLPQLLTLAEYHEQEBIFKLRLGH
LKKEEAEIQAELERLERVRNLHIRELKRINNEDNSQFKDHPTLNERYLLLHLLGRGGFSEVYKAFDLYEQRYAAVKIHQLNKSWRDEKKENYHKHACREYRIHKEL
DHPRIVKLYDYFSLDTDTFCTVLEYCEGNDLDFYLKQHKLMSEKEARSIVMQIVNALRYLNEIKPPIIHYDLKPGNILLVDGTACGEKITDFGLSKIMDDDSYGVDG
MDLTSQGAGTYWYLPPECFVVGKEPPKISNKVDVWSVGVIFFQCLYGRKPFGHINQSQQDILQENTILKATEVQFPVKPVVSSEAKAFIRRCLAYRKEDRFDVHQLA
NDPYLLPHMRRSNSSGNLHMAGLTASPTPSSSIITY

>SGK071_H_SEQID#AA_104
MLGPGSNRRRPTQGERGPGSPGEPMEKYQVLYQLNPGALGVNLVVEEMETKVKHVIKQVECMDDHYASQALEELMPLLKLRHAHISVYQELFITWNGEISSLYLC
LVMEFNELSFQEVIEDKRKAKKIIDSEWMQNVLGQVLDALEYLHHILDIHRNLKPSNILISSDHCKLQDLSSNVLMTDKAKWNIRAEEDPFRKSWMAPEALNFSFS
QKSDIWSLGCIILDMTSCSFMDGTEAMHLRKSLRQSPGSLKAVLKTMEEKQIPDVETFRNLLPLMLQIDPSDRTIKDVVHITFLRGSFKSSCVSLTLHRQMVPASITD
MLLEGNVASILGDAGDTKGERALKLLSMALASYCLVPEGSLFMPLALLHMHDQWLSCDQDRVPGKRDFASLGKLGKLLGPIPKGLPWPPELVEVVTTMELHDR
VLDVQLCACSLLLHLLGQGIIVNKAPLEKVPDLISQVLATYPADGEMAEASCGVFWLLSLLGCIKEQQFEQVVALLLQSIRLCQDRALLVNNAYRGLASLVKVSEL
AAFKVVVQBEGGSGLSLIKETYQLHRDDPEVVENVGMLLVHLASYEEILPELVSSSMKALLQEIKERFTSSLELVSCAEKVLLRLEAATSPSPLGGEAAQP

>SK516_H_SEQID#AA_105

MEGDGVPWGSEPVSGPGPGGGGMIRELCRGFGRYRYLGRLRQNLRETQKFFRDIKCSHNHTCLSSLTGGGAERGPAGDVAETGLQAGQLSCISFPPKEEKYLQ
QIVDCLPCILILGQDCNVKCQLLNLLGVQVLPTTKLGSEESCKLRRLRFTYGTQIRVSLALPGQYELVHTLVAHQGNWETIPEEDLEVQENNEDAAHVLAELEVT
MHHALLQEVDVVVAPCQGLRPTVDVLGDLVNDFLPVITYALHKDELSERDEQELQEIRKYFSFPVFFKVPKLGSEIIDSSTRRMESERSPLYRQLIDLGYLSSSHWN
CGAPGQDTKAQSMLVEQSEKLRHLSTFSHQVLQTRLVDAAKALNLVHCHCLDIFNQAFDMQRDLQITPKRLEYTRKKENELYESLMNIANRKQEEMKDMIVETL
NTMKEELLDDATNMEFKDVTVPENGEPVGTREIKCCIRQIQELIISRLNQAVANKLISSVDYLRESFVGTLERCLQSLEKSQDVSVHITSNYLKQILNAAYHVEVTFHS
GSSVTRMLWEQIKQIIQRITWVSPPAITLEWKRKVAQEAIESLSASKLAKSICSQFRTRLNSSHEAFAASLRQLEAGHSGRLEKTEDLWLRVRKDHAPRLARLSLESR
SLQDVLLHRKPKLGQELGRGQYGVVYLCDNWGGHFPCALKSVVPPDEKHWNDLALEFHYMRSLPKHERLVDLHGSVIDYNYGGGSSIAVLLIMERLHRDLYTGL
KAGLTLETRLQIALDVVEGIRFLHSQGLVHRDIKLKNVLLDKQNRAKITDLGFCKPEAMMSGSIVGTPIHMAPELFTGKYDNSVDVYAFGILFWYICSGSVKLPEAF
ERCASKDHLWNNVRRGARPERLPVFDEBCWQLMEACWDGDPLKRPLLGIVQPMLQGIMNRLCKSNSEQPNRGLDDST

>H85389_H_SEQID#AA_106
METGKDGARRGTQSPERKRRSPVPRAPSTKLRPAAAARAMDPVAAEAPGEAFLARRRPEGGGGSARPRYSLLAEIGRGSYGVVYEAVAGRSGARVAVKKIRCDA
PENVELALAEFWALTSLKRRHQNVVQFEECVLQRNGLAQRMSHGNKSSQLYLRIVETSLKGERILGYAEEPCYLWFVMEFCEGGDLNQYVLSRRPDPATNKSFM
LQLTSAIAFLHKNHVHRDLKPDNILITERSGTPILKVADFGLSKVCAGLAPRGKEGNQDKNVNVNKYWLSSACGSDFYMAPEVWEGHYTAKADIFALGIIWAM
IERITFIDSETKELLGTYIKQGTEIVPVGEALLENPKMELHIPQKRRTISMSEGIKQLLKDMLAANPQDRPDAFELETRMDQVTCAA

>Wee1b_H_SEQID#AA_107
MDDKDIDKELRQKLNFSYCEETEIEGQKKVEESREASSQTPEKGEVQDSEAKGTPPWTPLSNVHELDTSSEKDKESPDQILRTPVSHPLKCPETPAQPDSRSKLLPSD
SPSTPKTMLSRLVISPTGKLPSRGPKHLKLTPAPLKDEMTSLALVNINPFTPESYKKLFLQSGGKRKIRGDLEEAGPEEGKGGLPAKRCVLRETNMASRYEKEFLEVE
KIGVGEFGTVYKCIKRLDGCVYAIKRSMKTFTELSNENSALHEVYAHAVLGHHPHVVRYYSSWAEDDHMIIQNEYCNGGSLQAAISENTKSGNHFEEPKLKDILLQ
ISLGLNYIHNSSMVHLDIKPSNIFICHKMQSESSGVIEEVENEADWFLSANVMYKIGDLGHATSINKPKVEBGDSRFLANEILQEDYRHLPKADIFALGLTIAVAAGAE
SLPTNGAAWHHIRKGNFPDVPQELSESFSSLLKNMIQPDAEQRPSAAALARNTVLRPSLGKTEELQQQLNLEKFKTATLERELREAQQASPQGYTHHGDTGVSGT
HTGSRSTKRLVGGKSARSSSFTSGEREPLH

>Wnk2_H_SEQID#AA_108
MEPGRGAGPAGMAEPRAKAARPGPQRFLRRSVVESDQEEPPGLEAAEAPGPQPPQPLQRRVLLLCKTRRLIAERARGRPAAPAPAALVAQPGAPGAPADAGPEPV
GTQEPGPDPIAAAVETAPAPDGGPREEAAATVRKEDEGAAEAAKPEPGRTRRDEPEEEEDDEDDLKAVATSLDGRFLKFDIELGRGSFKTVYKGLDTETWVEVAWC
ELQDRKLTKLERQRFKEEAEMLKGLQHPNIVRFYDFWESSAKGKRCIVLVTELMTSGTLKTYLKRFKVMKPKVLRSWCRQILKGLLFLHTRTPPIIHRDLKCDNIFI
TGPTGSVKIGDLGLATLKRASFAKSVIGTPEFMAPEMYEEHYDESVDVYAFGMCMLEMATSEYPYSECQNAAQIYRKVTCGIKPASFEKVHDPEIKEIIGECICKNK
EERYEIKDLLSHARFAEDTGVRVELAEBDHGRKSTIALRLWVEDPKKLKGKPKDNGAIEFTFDLEKETPDEVAQEMESGFFHESDVKIVAKSIRDRVALIQWRRERI
WPALQPKEQQDVGSPDKARGPPVTLQVQVTYHAQAGQPGPPEPEPEADQHLLPPTLPTSAATSLASDSTFDSGQGSTVYSDSQSSQQSVMLGSLADAAPSPAQCVC
SPPVSEGPVLPQSLPSLGAYQQPTAAPGLPVGSVPAPACPPSLQQHFPDPAMSFAPVLPPPSTPMPTGPGQPAPPGQQPPLAQPTPLPQVLAPQPVVPLQPVPPHLPP
YLAPASQVGAPAQLKPLQMPQAPLQPLAQVPPQMPPIPVVPPITPLAGIDGLPPALPDLPTATVPPVPPQYFSPAVILPSLAAPLPPASPALPLQAVKLPHPGAPLA
MPCRTIVPNAPATPLLAVAPPGVAALSHSAVAQLPGQVPYPAAFPQMAPTDVPPSPHHTVQNMRATPPQPALPQPTLPPQVPPQPTLPPQVTRPPQPV
LPPQPMLPPQPVLPPQPALPVRPEPLQPHLPEQAAPAATPGSQILLGHPAPYAVDVAAQVPTVPVPPAAVLSPPLPEVLLPAAEPLLPQFPSSLATVSASVQSVPTQTA
TLLPPANPPLPGGPGIASPCPTVQLTVEPVQEEQASQDKPPGLPQSCESYGGSDVTSGKELSDSCEGAFGGRLEGRAARKHHRSTRARSRQERASRPRLIILNVCN
TGDKMVECQLETHNHKMVTFKFDLDGDAPDEIATYMVEHDFILQAERETFIEQMKDVMDKAEDMLSEDTDADRGSDPGTSPPHLSTCGLGTGEESRQSQANAPV
YQQNVLHTGKRWFICPVAEHPAPEAPESSPPLPLSSLPPEASQDSAPYKDQLSSKEQPSFLASQQLLSQAGPSNPPGAPPAPLAPSSPPVTALPQDGAAPATSTMPEP
ASGTASQAGGPGITPQGLTSELETSQPLAETHEAPLAVQPLVVGLAPCTPAPEAASTRDASAPREPLPPPAPEPSPHSGTPQPALGQPAPLLPAAVGAVSLATSQLPSPP

LGPTVPPQPPSALESDGEGPPPRVGFVDSTIKSLDEKLRTLLYQEHVPTSSASAGTPVEVGDRDFTLEPLRGDQPRSEVCGGDLALPPVPKEAVSGRVQLPQPLVEKS
ELAPTRGAVMEQGTSSSMTESSPRSMLGYDRDGRQVASDSHVVPSVPQDVPAFVRPARVEPTDRDGEAGESSAEPPSDMGTVGGQASHPQTLGARALGSPRKR
PEQQDVSSPAKTVGRFSVVSTQDEWTLASPHSLRYSAPPDVYLDEAPSSPDVKLAVRRAQTASSIEVGVGEPVSSDSGDEGPRARPPVQKQASLPVSGSVAGDFVK
KATAFLQRPSRAGSLGPETPSRVGMKVPTISVTSFHSQSSYISSDNDSELEDADIKKELQSQKQEIEALYRRLGKPLPPNVGFFHTAPPTGRRR
KTSKSKLKAGKLLNPLVRQLKVVASSTGHLADSSRGPPAKDPAQASVGLTADSTGLSGKAVQTQQPCSVRASLSSDICSGLASDGGARGQGWTVYHPTSERVTY
KSSSKPRARFLSGPVSVSIWSALKRLCLGKEHSSRSSTSSLAPGPEPGPQPALHVQAQVNNSNNKGTFTDDLHKLVDEWTSKTVGAAQLKPTLNQLKQTQKLQD
MEAQAGWAAPGEARAMTAPRAGVGMPRLPPAPGPLSTTVPGAAPTLSVPTPDGALGTARRNQVWFGLRVPPTACCGHSTQPRGGQRVGSKTASFAASDPVRS

>MAP3K1_H_SEQID#AA_109
MAAAAGNRASSSGFPGARATSPEAGGGGGALKASSAPAAAAGLLREAGSGGRERADWRRRQLRKVRSVELDQLPEQPLFLAASPPASSTSPSPEPADAAGSGTGF
QPVAVPPHGAASRRGAHLTESVAAPDSGASSPAAAEPGEKRAPAAEPSPAAAPAGREMENKETLKGLHKMDDRPEERMIREKLKATCMPAWKHEWLERRNRR
GPVVVKPIPVKGDGSEMNHLAEBSPGEVQASAASPASKGRRSPSPGNSPSGRITVKSESPGVRRKRVSPVPFQSGRITPPRRAPSPDGFSPYSPEETNRRVNKVMRAR
LYLLQQIGPNSFLIGGDSPDNKYRVFIGPQNCSCARGTFCIHLLFVMLRVFQLEPSDPMLWRKTLKNFEVESLFQKYHSRSRSRIKAPSRNTIQKFVSRMSNSHTLSSS
STSTSSSENSIKDEEEQMCPICLLGMLDEESLTVCEDGCRNKLHHHCMSIWAEECRRNREPLICPLCRSKWRSHDFYSHELSSPVDSPSSLRAAQQQTVQQPLAGSR
RNQESNFNLTHYGTQQIPPAYKDLABPWIQVFGMELVGCLFSRNWNVREMALRRLSHDVSGALLLANGESTGNSGGGSSGSSPSGGATSGSSQTSISGDVVEACCSV
LSMVCADPVYKVYVAALKTLRAMLVYTPCHSLAERIKLQRLLQPVVDTILVKCADANSRTSQLSISTLLEICKGQAGELAVGREILKAGSIGIGGVDYVLNCILGN
QTESNNWQELLGRLCLIDRLLLEFPABFYPHIVSTDVSQAEPVEIRYKKLLSLLTFALQSINNSHMVGKLSRRIYLSSARMVITVPHVFSKLLEMLSVSSTHFTRMR
RRLMAIADEVEIAEAIQLGVEDTLDGQQDSFLQASVPNNYLETTENSSPECTIHLEKTGKGLCATKLSASSEDISERLASISVGPSSSTTTTTEQPKPMVQTKGRPH
SQCLNSSPLSHHSQLMFPALSTPSSSTPSVPAGTATDVSKHRLQGFIPCRIPSASPQTQRKFSLQFHRNCPENKDSDKLSPVFTQSRPLPSSNIHRPKPSRPTPGNTSKQG
DPSKNSMTLDLNSSSKCDDSFGCSSNSSNAVIPSDETVFTPVEEKCRLDVNTELNSSIEDLLEASMPSSDTTVTFKSEVAVLSPEKAENDDTYKDDVNHNQKCKEKM
EAEEEEALAIAMAMSASQDALPIVPQLQVENGEDIIIQQDTPETLPGHTKAKQPYREDTEWLKGQQIGLGAFSSCYQAQDVGTGTLMAVKQVTYVRNTSSEQEEV
VEALREIRMMSHLNHPNIIRMLGATCEKSNYNLFIEWMAGGSVAHLLSKYGAFKESVVINYTEQLLRGLSYLHENQIIHRDVKGANLLIDSTGQRLRIADFGAAAR
LASKGTGAGEFQGQLLGTIAFMAPEVLRGQQYGRSCDVWSVGCAIIEMACAKPPWNAEKHSNHLALIFKIASATTAPSPSHLSPGLRDVALRCLELQPQDRPPSRE
LLKHPVFRITW

>MAP3K8_H_SEQID#AA_110
MPQIAKKQSTHRTQKPKKQSFPCICKNPGTQKSCVPLSVQPTEPRLNYLDLKYSDMFKEINSTANGPGIYEMFGTPVYCHVRETERDENTYYREICSAPSGRRITNKC
RSSHSERKSNIRTRLSQKKTHMKCPKTSFGIKQEHKVLISKEKSSKAVHSNLHDIENGDGISEPDWQIKSSGNEFLSSKDEIHPMNLAQTPEQSMKQNEFPPVSDLSIV
EEVSMEESTGDRDISNNQILTTSLRDLQELBELHHQIPFIPSEDSWAVPSEKNSNKYVQQEKQNTASLSKVNASRILTNDLEFDSVSDHSKTLTNFSFQAKQESASSQT
YQYWVHYLDHDSLANKSITYQMFGKTLSGTNSISQEIMDSVNNEELTDELLGCLAAELLALDEKDNNSCQKMANETDPENLNLVLRWRGSTPKEMGRETTKVKI
QRHSSGLRIYDREEKFLISNEKKIFSENSLKSEEPILWTKGEILGKGAYGTVYCGLTSQGQLIAVKQVALDTSNKLAAFKEYRKLQEEVDLLKALKHVNIVAYLGTC
LQENTVSIFMEFVPGGSISSIINRFGPLPEMVFCKYTKQILQGVAYLHENCVVHRDIKGNNVMLMPTGIIKLIDFGCARRLAWAGLNGTIHSDMLKSMHGTPYWMAP
EVINESGYGRKSDIWSIGCTVFEMATGKPPLASMDRMAAMFYIGAHRGLMPPLPDHFSENAADFVRMCLTRDQHERPSALQLLKHSFLERSH

>Pak5_M_SEQID#AA_111
MFGKKKKRVEISAPSNFEHRVHTGFDQHEQKFTGLPRQWQSLIEESARRPKPLIDPACITSIQPGAPKTIVRGSKGAKDGALTLLLDEFENMSVTRSNSLRRESPPPA
RAHQENGMLEERAAPARMAPDKAGSRARATGHSEAGSGSGDRRRVGPEKRPKSSRDGPGGPQEASRDKRPLSGPDVSTPQPGSLTSGTKLAAGRPFNTYRADTD
HPPRGAQGEPHTMAPNGPSATGLAAPQSSSSSRPPTRARGAPSPGVLGPHASEPQLAPPARALAAPAVPPAPGPPGPRSPQREPQRVSHEQFRAALQLVVDPGDPRS
YLDNFIKIGEGSTGIVCIATVRSSGKLVAVKKMDLRKQQRRELLFNEVVIMRDYRHENVVEMYNSYLVGDELWVVMEFLEGALITDIVTHTRMNEEQIAAVCLAV

LQALAVLHAQGVIHRDIKSDSILLTHDGRVKLSDFGFCAQVSKEVPRRKSLVGTPYWMAPELISRLPYGPEVDIWSLGVMVIEMVDGEPPYFNEPPLKAMKMIRDN
LPPRLKNLHKASPSLKGFLDRLLVRDPAQRATAAELLKHPFLTKAGPPASIVPLMRQHRTR

>STLK6-rs_H_SEQID#AA_112
MSLLDCFCTSRTQDESLRPEKQSETSIHQYLVDEPTLSWSPPSTRASEVLCSTNVSHYELQVETGRGFDNLTSVHLAWHTPTGTLVTIKITNLENCNEERLKALQKA
VILSHFFRRHPNITTYWTVFTVGSWLWVISPFMAYGSASQLLRTCFPEGMSETLIRNIIFGAVRGLNYLYQNGCIHRSIKASHILISGDGLVTLSGLSHLHSLVKHGQRH
RAVYDFPQFSTSVQPWLSPELLRQDLHGYNVKSDIYSVGITACELASGQVPFQDMHRTQMLLQKIKGPPYSPLDISIFPQSESRMKNSRSGVDSGIGESVLVSSGTIHT
VNSDRLHTPSSKTFSPAFFSLVQLCLQQDPEKRPSASSLLSHVFFKQMKBESQDSILSRLPPAYNKPSIPLPPVLPWTEPECDFPDEKDSYLEF

>MAP2K2_H_SEQID#AA_113
MLARRKPMLPALTINPTIAEGPSPTSEGASEANLVDLQNKLEELELDEQQKRLEAFLTQKAKVGELKDDDFERTSELDAGNGGVVTKVQHRPSGLIMARKLIHLEIK
PAIRNQIIREHQVLHECNSPYIVGFYGAFYCDREISICMEHMDGGSLDQGLKEAKRIPEDILGKVSIAVLRGLAYLREKHQIMHRNVKPSNILVNSRGEIKLCDFGVSG
QLIDSMANSFVGTRSYMAPERLQGTHYSVQSVIWSMDLSLVELAIERYPIPPDAKELEAIFGRPVVDREEGEPHSISSWPGSPGRPNSGHGMDSRPAMAIFELLDYI
VKEPPPKLPNGVFTPDFQEFVNKCLIKNPTERADLKMLTNHAFIKRSEVKEADFAC

>CCK4_H_SEQID#AA_114
MGAARGSPARPRRLPLLSVLLIPLLGGTQTAFVFIKQPSSQDALQGRRALLRCEVEAPGPVHVYWLLDGAPVQDTERRFAQGSSLSFAAVDRLQDSGTFQCVARDD
VTGEEAARSANASFNIKWIEAGPVVKHPASEAEIQPQTQVTLRCHIDGHPRPTYQWFRDGTPLSDGQSNHTVSSKERNLTLRPAGPEHSGLYSCCAHSAFGQACSSQ
NFTLSIADESFARVVLAPQDVVVARYEEAMFHCQFSAQPPPSLQWLFEDETPITNRSRPPHLRRATVFANGSLLLTQVRPRNAGIYRCIGQGQRGPPILEATLIHLAEI
EDMPLFEPRVFTAGSEERVTICLPPKGLPEPSVWWEHAGVRLPTHGRVYQKGHELVLANLAESDAGVYTCHAANLAGQRRQDVNITVATVPSWLKKPQDSQLEEG
KPGYLDCLTQATPKPTVVWYRNQMLISEDSRFEVFKNGTLRINSVEVYDGTWYRCMSSTPAGSIEAQARVQVLEKLKFTPPPQPQQCMEFDKEATVPCSATGREKP
TIKWERADGSSLPEWVTDNAGTLHFARVTRDDAGNYTCIASNGPQGQIRAHVQLTVAVFIIFKVEPERITVYQGHTALLQCEAQGDPKPLIQWKGKDRILDPTKL
GPRMHIFQNGSLVIHDVAPEDSGRYTCIAGNSCNIKHTEAPLYVVDKPVPEESEGPGSPPPYKMIQTIGLSVGAAVAYIIAVLGLMFYCKKRCKAKRLQKQPEGEEP
EMECLNGGPLQNGQPSAEIQEEVALTSLGSGPAATNKRHSTSDKMHFPRSSLQPITTLGKSEFGEVFLAKAQGLEEGVAETLVLVKSLQSKDEQQQLDFRRELEMF
GKLNHANVVRLLGLCREAEPHYMVLEYVDLGDLKQFLRISKSKDEKLKSQPLSTKQKVALCTQVALGMEHLSNNRFVHKDLAARNCLVSAQRQVKVSALGLSK
DVYNSEYYHFRQAWVPLRWMSPEAILEGDFSTKSDVWAFGVLMWEVFTHGEMPHGGQADDEVLADLQAGKARLPQPEGCPSKLYRLMQRCWALSPKDRPSFSE
IASALGDSTVDSKP

>LMR1_H_SEQID#AA_115
MSSSFFNPSFAFSSHFDPDGAPLSELSWPSSLAVVAVSFSGLFAVIVLMLACLCCKKGGIGFKEFENAEGDEYAADLAQGSPATAAQNGPDVVYLPLTEVSLPMAK
QPGRSVQLLKSTDVGRHSLLYLKEIGRGWFGKVFLGEVNSGISSAQVVVKELQASASVQEQMQFLEEVQPYRALKHSNLLQCLAQCAEVTPYLLVMEFCPLGDLK
GYLRSCRVAESMAPDRRTLQRMACEVACGVLHLHRNNFVHSDLALRNCLLTADLTVKIGDYGLAHCKYREDYFVTADQLWVPLRWIAPELVDEVHSNLLVVDQ
TKSGNVWSLGVTIWELFELGTQPYPQHSDQQVLAYTVREQQLKLPKPQLQITLSDRWYEVMQFCWLQPEQRPTAEEVHILLSYLCAKGATEAEEFERRWRSLR
PGGGGVGPGPGAAGPMLGGVVELAAASSFPLLEQFAGDGFHADGDDVLTVTETSRGLNFEYKWEAGRGAEAFFATLSPGRTARLQELCAPDGAPPGVVPVLSAH
SPSLGSEYFIRLEEAAPAAAGHIDPDCAGCAPSPPATADQDDDSDGSTAASLAMEPLLGHGPPVDVPWGRGDHYPRRSLARDPLCPSRSPSAGPLSLAEGGAEDAD
WGVAAFCPAFFEDPLGTSPLGSSGAPPLPLTGEDELEEVGARRAAQRGHWRSNVSANNNSGSRCPESWDPVSAGCHAEGCPSPKQTPRASPEPGYPGEPLLGLQAA
SAQEPGCCPGLPHLCSAQGLAPAPCLVTPSWTETASSGDHPQAEPKLATEAEGTTGPRLPLPSVPSPSQEGAPLPSEBASAPDAPDALPDSPTPATGGEVSAIKLAS
ALNGSSSSPEVEAPSSEDEDTAEATSGIFIDTSSDGLQARRPDVVPAFRSLQKQVGTPDSLDLDIPSSASDGGYEVFSPSATGPSGQPRALDSGYDTENYESPEFVL
KEAQEGCBPQAFABLASEGEGPGPETRLRSTSLSGLNEKNPYRDSAYFSDLEAEAEATSGPEKKCGGDRAPGPELGLPSTGQPSEQVCLRPGVSGEAQGSGPGEVLPP

LLQLEGSSPEPSTCPSGLVPEPPEPQGPAKVRPGPSPSCSQFFLLTPVLRSEGNSSEFQGPPGLLSGPAPQKRMGGPGTPRAPLRLALPGLPAALEGRPEEEEDSEDS
DESDEELRCYSVQEPSEDSEEEAPAVPVVAESQSARNLRSLLKMPSLLSETFCEDLERKKKAVSFFDDVTVYLFDQESPTRELGEPFFPGAKESPPTFLRGSPGSPSAP
NRPQQADGSPNGSTAEEGGGFAWDDIFPLMTAKAAFAMALDPAAPAPAAPTPTPAPFSRFTVSPAPTSRFSITHVSDSDAESKRGPEAGAGGESKEA

>RYK_H_SEQID#AA_116
MRGAARLGRPGRSCLPGARGLRAPPPPPLLLLLALLPLLPAPGAAAAPARPPELQSASAGPSVSLYLSEDEVRRLIGLDAELYYVRNDLISHYALSFSLLVPSETNFL
HFTWHAKSKVEYKLGFQVDNVLAMDMPQVNISVQGEVPRTLSVFRVELSCTGKVDSEVMILMQLNLTVNSSKNFTVLNFKRRKMCYKKLEVKTSALDKNTSRT
IYDPVHAAPTTSTRVFYISVGVCCAVIFLVAIILAVLHLHSMKRIELDDSISASSSSQGLSQPSTQTTQYLRADTPNNATPITSYPTLRIEKNDLRSVTLLEAKGKVKDI
AISRERITLKDVLQEGITFGRIFHGILIDEKDPNKEKQAFVKTVKDQASEIQVTMMLTESCKLRGLHHRNLLPITHVCIEEGEKPMVILPYMNWGNLKLFLRQCKLVE
ANNPQAISQQDLVHMAIQIACGMSYLARREVIHKDLAARNCVIDDTLQVKITDNALSRDLFPMDYHCLGDNENRPVRWMALESLVNNEFSSASDVWAFGVTLWE
LMTLGQITPYVDIDPFEMAAYLKDGYRIAQPINCPDELFAVMACCWALDPEERPKFQQLVQCLTEFHAALGAYV

>LRRK2_H_SEQID#AA_117
MASGSCQGCEEDEETLKKLLIVRLNNVQEGKQIETLVQILEDLLVFTYSEHASKLFQGKNIHVPLLIVLDSYMRVASVQQVGWSLLCKLIEVCPGTMQSLMGPQDVG
NDWEVLGVHQLILKMLTVHNASVNLSVIGLKTLDLLTSGKITLLILDEESDIFMLIFDAMHSFPANDEVQKLGCKALHVILFERVSEEQLTEFVENKDYMILLSALT
NFKDEEEIVLHVLHCLHSLAIPCNNVBVLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVVKAVQQYPENAALQISALSCLALLT
ETIFLNQDLEEKNENQENDDEGEHEDKLFWLEACYKALTWHRKNKHVQEAACWALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKEVFQASANALS
TLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEGSNTSLDIMAAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDTEFHHKLN
MVKKQCFKNDIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEMLSLEGAMDSVLHTLQMYPDDQEIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFK
DVAEIQTKGFQTILAILKLSASFSKLLVHHSFDLVIFHQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYLKNVMLERACDQNNSIMVECLLLGADANQAKEGSSL
CQVCEKESSPKLVELLLNSGSREQDVRKALTISIGKGDSQIISLLLRRLALDVANNSICLGGFCIGKVEPSWLGPLFPDKTSNLRKQTNLASTLARMVIRYQMKSAVEE
GTASGSDGNFSEDVLSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEGSFLVKKKSNSISVGEFYRDAVLQRCSPNLQRHSNSLGPIFDHEDLLKRKRKILSSDDSLRSS
KLQSHMRHSDSISSLASEREYITSLDLSANELRDIDALSQKCCISVHLEHLEKLELHQNALTSFPQQLCETLKSLTHLDLHSNKFTSPPSYLLKMSCIANLDVSRNDIGP
SVVLDPTVKCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNKISGICSPLRLKELKILNLSKNHISSLSENFLEACPKVESFSARMNFLAAMPFLPPSMTILKLSQ
NKFSCIPEAILNLPHLRSLDMSSNDIQYLPGPAHWKSLNLRELLFSHNQISILDLSEKAYLWSRVEKLHLSHNKLKEIPPEIGCLENLTSLDVSYNLELRSFPNEMGKLS
KIWDLPLDELHLNFDFKHIGCKAKDIRFLQQRLKKAVPYNRMKLMIVGNTGSGKTTLLQQLMKTKKSDLGMQSATVGIDVKDWPIQIRDKRKRDLVLNVWDFA
GREEFYSTHPHFMTQRALYLAVYDLSKGQAEVDAMKPWLFNIKARASSSPVILVGTHLDVSDEKQRKACMSKITKELLNKRGFPAIRDYHFVNATEESDALAKLR
KTIINESLNFKIRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKRLLQLVRENQLQLDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQD
VSSIFGLYTRDILTVKVEGCPKHPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLSDHRPVIELPHCENSEIIRLYEMPYFPMGFWSRLIN
RLLEISPYMLSGRGCILLGQVVDHIDSLMBEWFPGLLEIDICGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLILADLPRNI
MLNNDELEFEQAPEFLLDCFVCIHLYPSSDYISRHYMRTINIVQTGFAKCRWRVTVHGADHGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQELVVLCHLHHP
SLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVADGLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTPGF
RAPEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQQGKLPDPVKEYGCAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILL
PKNVIVBCMVATHHNSRNASIWLGCCHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKBSWIVSGTQSGTLLVINTEDGKKRHTLEKMTDSVTCLYCNS
FSKQSKQKNFLLVGTADGKLAIFEDKTVKLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRTSQLFSYAAFSDSNIITVVDT
ALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKENKESKHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVTYNFCNSVRVMMTAQLGSL
KNVMLVLGYNRKNTBGTQKQKBIQSCLTVWDNLPHEVQNLEKHIEVRKELABKMRRTSVE

>pMLK4_H_SEQID#AA_118

MALRGAAGATDTPVSSAGGAPGGSASSSTSSGGSASAGAGLWAALYDYEARGEDELSLRRGQLVEVLSQDAAVSGDEGWWAGQVQRRLGIFPANYVAPCRPA
ASPAPPPSRPSSPVHVAFERLELKELIGAGGFGQVYRATWQGEVAVKAARQDPEQDAAAAAESVRREARLFAMLRHPNIELRGVCLQQPHLCLVLEFARGGAL
NRALAAANAAPDRAPGPRRARRIPPHVLVNWAVQIARGMLYHEEAFVPILHRDLKSSNILLEKIEHDDICNKTLKITDFGLAREWHRTTKMSTAGTYAWMAPE
VIKSSLFSKGSDIWSYGVLLWELLTGEVPYRGIDGLAVAYGVAVNKLTLPIPSTCPEPFAKLMKECWQQDPHIRPSFALILEQLTAIEGAVMTEMPQESFHSMQDDW
KLEIQQMFDELRTKEKELRSREELTRAALQQKSQEBLLKRREQQLAEREIDVLERELNLIFQLNQEKPKVKKRKGKFKRSRLKLKDGHRISLPSDFQHKITVQASP
NLDKRRSLNSSSSSPPSSPTMMPRLRAIQLTSDESNKTWGRNTVFRQEEFEDVKRNFKKKGCTWGPNSIQMKDRTDCKERIRPLSDGNSPWSTILKNQKTMPLASL
FVDQPGSCEEPKLSPDGLEHRKPKQIKLPSQAYIDLPLGKDAQRENPAEAESWEEAASANAATVSIEMTPTNSLSRSPQRKKTESALYGCTVLLASVALGLDLRELH
KAQAAEEPLPKEEKKKREGIFQRASKSRRSASPPTSLPSTCGEASSPPSLPLSSALGILSTPSFSTKCLLQMDSEDPLVDSAPVTCDSEMLTPDFCPTAPGSGREPALMP
RLDTDCSVSRNLPSSFLQQTCGNVPYCASSKHRPSHHRRTMSDGNPTPTGATIISATGASALPLCPSPAPHSHLPREVSPKKHSTVHIVPQRRPASLRSRSDLPQAYPQ
TAVSQLAQTACVVGRPGPHPTQFLAAKERITKSHVPSLLDADVEGQSRDYTVPLCRMRSKTSRPSIYELEKEFLS

>KSR_H_SEQID#AA_119
MDRAALRAAAMGEKKEGGGGGDAAAAEBGAGAAASRALQQCGQLQKLIDISIGSLRGLRTKCAVSNDLTQQERTLEAKLVRYICKQRQCKLSVAPGERTPELNS
YPRFSDWLYTFNVRPEVVQEIPRDLTLDALLEMNEAKVKETLRRCGASGDECGRLQYALTCLRKVTGLGGEHKEDSSWSSLDARRESGSGPSTDTLSAASLPWPP
GSSQLGRAGNSAQGPRSISVSALPASDSPTPSFSEGLSDTCIPLHASGRLTPRALHSFITPPTTPQLRRHTKLKPRRTPPPPSRKVFQLLPSFPTLTRSKSHESQLGNRIDD
VSSMRFDLSHGSPQMVRRDIGLSVTHRFSTKSWLSQVCHVCQKSMIFGVKCKHCRLKCHNKCTKEAPACRISFLPLTRLRRTESVPSDINNPVDRAAEPHFGTLPKA
LTKKEHPPAMNHLDSSNPSSTTSSTPSSPAPFTSSNPSSATTPPNPSPGQRDSRFNFPAAYFIHHRQQFIFPDISAFAHAAPLPEAADGTRLDDQPKADVLEAHEAEA
EEPEAGKSEAEDDEDEVDDLPSSRRPWRGPISRKASQTSVYLQEWDIPFEQVELGEPIGQGRWGRVHRGRWHGEVAIRLLEMDGHNQDHLKLFKKEVMNYRQTR
HENVVLFMGACMNPPHLAIITSFCKGRTLHSFVRDPKTSLDINKTRQIAQEIIKGMGYLHAKGIVHKDLKSKNVFYDNGKVVITDFGLFGISGVVREGRRENQLKLS
HDWLCYLAPEIVREMTPGKDEDQLPFSKAADVYAFGTVWTELQARDWPLKNQAAEEASIWQIGSGEGMKRVLTSVSLGKEVSEILSACWAFDLQERPSFSLLMDM
LEKLPKLNRRLSHPGHFWKSADINSSKVVPRFERRGLGVLESSNPKM

>KSR2_H_SEQID#AA_120
MGAAGGMQASSLTRAQLAAGKEEEGKGVAEQAMDEENMTKSEEQQPLSLQKALQQCELVQNMIDLSISNLEGLRTKCATSNDLTQKEIRTLESKLVKYFSRQLSC
KKKVALQERNAELDGFPQLRHWFRIVDVRKEVLEEISPGQLSLEDLLEMTDEQVCETVEKYGANREECARLNASLSCLRNVHMSGGNLSKQDWTIQWPTETGKE
NNPVCPPEPTPWIRTHLSQSPRVPSKCVQHYCHTSPTPGAPVYTHVDRLTVDAYPGLCPPPPLESGHRSLPPSPRQRHAVRTPPRTPNIVTIVTPPGTPPMRKKNKLK
PPGTPPPSSRKLIHLIPGFTALHRSKSHEFQLGHRVDEAHTPKAKKKSKPLNLKIHSSVGSCENIPSQQRSPLLSERSLRSFFVGHAPFLPSTPPVHTEANFSANTLSVPR
WSPQIPRRDLGNSIKHRSTKYWMSQTCTVCGKGMLFGLKCKNCKLKCHNKCTKEAPPCHLIIHRGDPARLVRTESVPCDINNPLRKPPRYSDLHISQTLPKTNKI
NKDHIPVPYQPDSSSNPSSTTSSTPSSPAPPLPPSATPPSPLHPSPQCTRQQKNFNLPASHYYKYKQQFIFPDVVPVPETPTRAPQVILHPVTSNPLLEGNPLLQIEVEPTS
ENEEVHDEAEESEDDFEEMNLSLLSARSFPRKASQTSIFLQEWDPFEQLEIGELIGKGRFGQVYHGRWHGEVAIRLLDIERDNEDQLKAFKREVMAYRQTRHENVV
LFMGACMSPPHLAIITSLCKGRTLYSVVRDAKIVLDVNKTRQIAQEIVKGMGYLHAKGLHKDLKSKNVFYDNGKVVITDFGLFSISGVLQAGRREDKLRIQNGWL
CHLAPEIIRQLSPDTEEDKLPFSKHSDVFALGTIWYELHAREWPFKTQPAEAIIWQMGTGMKPNLSQIGMGKEISDILLFCWAFEQERPTFTKLMDMLEKLPKRNR
RLSHPGHFWKSAEL

>KIAA1646_H_SEQID#AA_121
MGATGAAEPLQSVLWVKQQRCAVSLEPARALLRWWRSPGPGAGAPGADACSVPVSEIIAVEETDVHGKHQGSGKWQKMEKPYAFTVHCVKRARRHRWKWAQ
VTFWCPEEQLCHLWLQTLREMLEKLTSRPKHLLVFINPFGGKGQGKRIYERKVAPLFTLASITDIVTEHANQAKETLYEINDKYDGIVCVGGDGMFSEVLHGLIG
RTQRSAGVDQNHPRAVLVPSSLRIGIIPAGSTDCVCYSTVGTSDAETSALHIVVGDSLAMDVSSVHHNSTLRYSVSLLGYFYGDIIKDSEKKRWLGLARYDFSGL
KITFLSHHCYEGTVSFLPAQHTVGSPRDRKPCRAGCFVCRQSKQQLEEEQKALYGLEAAEDVEEWQVVCGKFLAINATNMSCACRRSPRGLSPAAHLGDGSSDLI

LIRKCSRFNFLRFLIRHTNQQDQFDFTFVEVYRVKKFQFTSKHMEDEDSDLKEGGKKRFGHICSSHPSCCCTVSNSSWNCDGEVLHSPAIEVRVHCQLVRLFARGIE
ENPKPDSHS

>DGK-beta_H_SEQID#AA_122
MTNQEKWAHLSPSEFSQLQKYAEYSTKKLKDVLEEFHGNGVLAKYNPEGKQDILNQTIDFEGFKLFMKTFLEAELPDDFTAHLFMSFSNKFPHSSPMVKSKPALLS
GGLRMNKGAITPPRITSPANTCSPEVIHLKDIVCYLSLLERGRPEDKLEFMFRLYDTDGNGFLDSSELENIISQMMHVAEYLEWDVTELNPILHEMMEEIDYDHDGT
CPTKCDKCHKTVKCYQGLTGLHCVWCQITLHNKCASHLKPECDCGPLKDHILPPTTICPVVLQTLPTSGVSVPEERQSTVKKEKSGSQQPNKVIDKNKMQRANSVT
VDGQGLQVTPVPGTHPLLVFVNPKSGGKQGERIYRKFQYLLNPRQVYSLSGNGPMPGLNFFRDVPDFRVLACGGDGTVGWVLDCIEKANVGKHPPVAILPLGTGN
DLARCLRWGGGYEGENLMKILKDIENSTEIMLDRWKFEVIPNDKDEKGDGPVPYSIINNYFSIGVDASIAHRFHIMREKIHPEKFNSRMKCNKFWYFEFGTSETFSATCK
KLHESVEIECDGVQIDLNISLEGIAILNIPSMHGGSNLWGESKKRRSHRRIEKKGSDKRTTVTDAKELKFASQDLSDQLLEVVGLEGAMEMGQIYTGLKSAGRRLA
QCSCVVIRTSKSLPMQIDGEPWMQTPCTIKITHKNQAPMLMGPPPKTGLFCSLVKRTRNRSKE

>P6K1_H_SEQID#AA_123
MCVCQTMEVGQYGKNASRAGDRGVLLEPFIHQVGGHSSMMRYDDHTVCKPLISREQRFYESLPPEMKEFTPEYKGVVSVCFEGDSDGYINLVAYPYVESETVEQ
DDTTEREQPRRKHSRRSLHRSGSGSDHKEEKASLSLETSESSQEAKSPKVELHSHSEVPFQMLDGNSGLSSEKISHNPWSLRCHKQQLSRMRSESKDRKLYKFLLE
NVVHHFKYPCVLDLKMGTRQHGDDASAEKAARQMRKCEQSTSATLGVRVCGMQVYQLDTGHYLCRNKYYGRGLSIEGFRNALYQYLHNGLDLRRDLFEPILSK
LRGLKAVLERQASYRFYSSSLLVIYDGKECRAESCLDRRSEMRLKHLDMVLPEVASSCGPSTSPSNTSPEAGPSSQPKVDVRMIDFAHSTFKGFRDDPTVHDGPDRG
YVFGLENLISIMEQMRDENQ

>YAB1_H_SEQID#AA_124
MAAILGDTIMVAKGLVKLTQAAVETHLQHLGIGGELIMAARALQSTAVEQIGMFLGKVQGQDKHEEYFAENFGGPEGEFHFSVPHAAGASTDFSSASAPDQSAPPS
LGHAHSEGPAPAYVASGPFREAGFPGQASSPLGRANGRLFANPRDSFSAMGFQRRFHQDQSPVGGLTAEDIEKARQAKARPENKQHKQTLSEHARERKVPVTRIG
RLANFGGLAVGLGFGALAEVAKKSLRSEDPSGKKAVLGSSPFLSEANAERIVRTLCKVRGAALKLGQMLSIQDDAFINPHLAKIFERVRQSADFMPLKQMMKTLN
NDLGPNWRDKLEYFEERPFAAASIGQVHLARMKGGREVAMKIQYPGVAQSINSDVNNLMAVLNMSNMLPEGLFPEHLIDVLRRELALECDYQREAACARKFRDL
LKGHPFFYVPEIVDELCSPHVLTTELVSGFPLDQAEGLSQEIRNEICYNILVLCLRELFEFHFMQTDPNWSNFFYDPQQHKVALLDFGATREYDRSFTDLYIQHRAAA
DRDRETVRAKSIEMKFLTGYEVKVMEDAHLDAILILGEAFASDEPFDFGTQSTTEKIHNLIPVMLRHRLVPPEETYSLHRKMGGSFLICSKLKARFPCKAMFEEAYS
NYCKRQAQQ

>AF052122_H_SEQID#AA_125
GRGEAAAAPLNPGGGSHWLGLRLRLRGNDEAGHRAQGATLQVRVQAMWLKVGGLLRGTGGQLGQTVGWPCGALGPGPHRWGPCGGSWAQKFYQDGPRG
LGEEDIRRAREARPRKTPRPQLSDRSRERKVPASRISRLANFGGLAVGLGVLAEMAKKSMPGGRLQSEGGSGLDSSPFLSEANAERIVQTLCTVRGAALKVGQM
LSIQDNSFISPQLQRIFERVRQSADFMPRWQMLRVLEEELGRDWQAKVASLEEVPFAAASIGQVHQGLLRDGTEVAVKIQYPGIAQSIQSDVQNLLAVLKMSAALP
AGLFAEQSLQALQQELAWECDYRREAACAQNFRQLLANDPFFRVPAVVKELCTTRVLGMELAGGVPLDQCQGLSQDLRNQICFQLLTLCLRELFEFRFMQTDPN
WANFLYDASSHQVTLLDFGASREFGTEFTDHYIEVVKAAADGRDCVLQKSRDLKFLTGFETKAFSDAHVEAVMILGEPFATQGPYDFGSGETARRIQDLIPVLLR
HRLCPPPEETYALHRKLAGAFLACAHLRAHIACRDLFQDTYHRYWASRQPDAATAGSLPTKGDSWVDPS

>AAF23326_H_SEQID#AA_126

MARKALKLASWTSMALAASGIYFYSNKYLDPNDFGAVRVGRAVATTAVISYDYLTSLKSVPYGSEEYLQLRSKIHDLFQSFDDTPLGTASLAQVHKAVLHDGRTV
AVKVQHPKVRAQSSKDILLMEVLVLAVKQLFPEFEFMWLVDEAKKNLPLELDFLNEGRNAEKVSQMLRHFDFLKVPRIHWDLSTERVLLMEFVDGGQVNDRDY
MERNKIDVNEISRHLGKMYSEMIFVNGFVHCDPHPGNVLVRKHPGTGKAEIVLLDHGLYQMLTEEFRLNYCHLWQSLIWTDMKRVKEYSQRLGAGDLYPLFACM
LTARSWDSVNRGISQAPVTATEDLEIRNNAANYLPQISHLLNHVPRQMLLILKTNDLLRGIEAALGTRASASSFLNMSRCCIRALAEHKKKNTCSFFRRTQISFSEAF
NLWQINLHELILRVKGLKLADRVLALICWLFPAPL

>SGK493_H_SEQID#AA_127
MGKVNVAKLRYMSRDDFRVLTAVEMGMKNHEIVPGSLIASIASLKHGGCNKVLRELVKHKLIAWERTKTVQGYRLTNAGYDYLALKTLSSRQVVESVGNQMGV
GKESDIYIVANEEGQQFALKLHRLGRTSFRNLKNKRDYHKRHNVSWLYLSRLSAMKEFAYMKALYERKFPVPKPIDYNRHAVVMELINGYPLCQHHVEDPASV
YDEAMELIVKLANHGLIHGDFNEFNLILDESDHITMIDFPQMVSTSHPNAEWYFDRDVKCIKDFFMKRFSYESELFPTFKDIRREDTLDVEVSASGYTKEMQADDEL
LHPLGPDDKNETKEGSEFSFSDGEVAEKAEVYRSENESERNCLEESEGCYCRSSGDPEQIKEDSLSEESADARSFEMTEFNQALEEKGQVVENNSVTEFSEEKNRI
ENYNRQDGQRVQGGVPAGSDEYEDECPHLIALSSLNREFRPFRDEENVGAMNQYRTRTLSTTSSGSAVSCSTTPPELVKQKVKRQLTKQQKSAVRRRLQKGEANIFT
KQRRENMQNIKSSLEAASFWGE

>BRD2_H_SEQID#AA_128
MLQNVTPHNKLPGEGNAGLLGLGPEAAAPGKRIRKPSLLYEGFESPTMASVPALQLTPANPPPEVSNPKKPGRVTNQLQYLHKVVMKALWKHQFAWPFRQPVD
AVKLGLPDYHKIIKQPMDMGTIKKRLENNYYWAASECMQDFNTMFTNCYTNKPTDDIVLMAQTLEKIFLQKVASMPQEEQELVVTIPKNSHKKGAKLAALQGS
VTSAHQVPAVSSVSHTALYTPPPEIPTVLNIPHPSVISSPLLKSLHSAGPPLLAVTAAPPAQPLAKKKGVKRKADTTTPTTAILAPGSPASPPGSLEPKAARLPPMRR
ESGRPIKPPRKDLPDSQQHQSSKKGKLSEQLKHCNGILKELLSKKHAAYAWPFYKPVDASALGLHDYHDIIKHPMDLSTVKRKMENRDYRDAQEFAADVRLMFS
NCYKYNPPDHDVVAMARKLQDVFEFRYAKMPDEPLEPGPLPVSTAMPPGLAKSSSSESSSESSSSEEEEEDEEDEEEESESSDSEEERAHRLAELQEQLRAVH
EQLAALSQGPISKPKRKREKKEKKKRKAEKHRGRAGADEDDKGPRAPRPPQPKKSKKASGSGGGSAAALGPSGFGPSGSGGSGTKLPKKATKTAPPALPTGYDSEEE
BESRPMSYDEKRQLSLDINKLPGEKLGRVVHIIQAREPSLRDSNPEEIEIDFETLKPSTLRELERYVLSCLRKKPRKPYTIKKPVGKTKEELALEKKRELEKRLQDVSG
QLNSTIKKPPKKANEKTESSSAQQVAVSRLSASSSSSDSSSSSSSSSDTSDSDSG

>BRD3_H_SEQID#AA_129
MSTATTVAPAGIPATPGPVNPPPEVSNPSKPGRKTNQLQYMQNVVVKTLWKHQFAWPFYQPVDAIKLNLPDYHKIIKNPMDMGTIKKRLENNYYWSASECMQD
FNTMFTNCYTNKPTDDIVLMAQALEKIFLQKVAQMPQEEVELLPPAPKGKGRKPAAGAQSAGTQQVAAVSSVSPATPFQSVPPTVSQTPVIAATPVPTITANVTSV
PVPPAAAPPPATPIVPVVPPTPPVKKKGVKRKADTTTPTTSAITASRSESPPPLSDPKQAKVARRESGGRPIKPPKKDLEDGEVPQHAGKKGKLSEHLRYCDSILR
EMLSKKHAAYAWPFYKPVDAEALELHDYHDIIKHPMDLSTVKRKMDGREYPDAQGFAADVRLMFSNCYKYNPPDHEVVAMARKLQDVFEMRFAKMPDEPVEA
PALPAPAAPMVSKGAESSRSSEESSDSGSSDSEEERATRLAELQEQLKAVHEQLAALSQAPVNKPKKKEKKEKKKDKEKEKEKHKVKAEEEKKAKVAPPA
KQAQQKKAPAKKANSTTTAGRQLKKGGKQASASYDSEEEEEGLPMSYDEKRQLSLDINRLPGEKLGRVVHIIQSREPSLRDSNPDEIEIDFETLKPTTLRELERYVKS
CLQKKQRKPFSASGKKQAAKSKEELAQEKKKELEKRLQDVSGQLSSKKPARKEKPGSAPSGGPSRLSSSSSESGSSSSSGSSSDSDSE

>BRD4_H_SEQID#AA_130
MSAESGPGTRLRNLPVMGDGLETSQMSTTQAQAQPQPANAASTNPPPETSNPNKPKRQTNQLQYLLRVVLKTLWKHQFAWPFQQPVDAVKLNLPDYYKIIKTP
MDMGTIKKRLENNYYWNAQECIQDFNTMFTNCYIYNKPGDDIVLMAEALEKLFLQKINELPTEETEIMIVQAKGRGRKETGTAKPGVSTVPNTTQASTPPQTQT
PQPNPPPVQATPHPFPAVTPDLIVQTPVMTVVPPQPLQTPPVPPQPQPPPAPAPQPVQSHPIIAATPQPVKTKKGVKRKADTTTPTTIDPIHEPPSLPPEPKTTKLGQR
RESSRPVKPPKKDVPDSQQHPAPEKSSKVSEQLKCCSGILKEMFAKKHAAYAWPFYKPVDVEALGLHDYCDIIKHPMDMSTIKSKLEAREYRDAQEFGADVRLMF
SNCYKYNPPDHEVVAMARKLQDVFEMRFAKMPDEPEEPVVAVSSPAVPPPTKVVAPPSSSDSSSSDSSDSSTDDSEEERAQRLAELQEQLKAVHEQLAALSQPQ

QNKPKKKEKDKKEKKEKHKRKEEVEENKKSKAKEPPPKKTKKNNSSNSNVSKKEPAPMKSKPPPTYESEEEDKCKPMSYEEKRQLSLDINKLPGEKLGRVVHIIQ
SREPSLKNSNPDEIEIDFETLKPSTLRELERYVTSCLRKKRKFQQAEKVDVLAGSSKMKGFSSSESESSSESSSSDSEDSETGPA

>BRDT_H_SEQID#AA_131
MSLPSRQTAIIVNPPPEYINTKKKNGRLTNQLQYLQKVVLKDLWKHSFSWPFQRPVDAVKLKLPDYYTIIKNPMDLNTIKKRLENKYYAKASECIEDFNTMFSNCYL
YNKPGDDIVLMAQALEKLFMQKLSQMPQEBQVVGVKERIKKGTQQNLAVSSAKEKSSPSATEKVFKQQEIPSVFPKTSISPLNVVQGASVNSSSQTAAQVTKGVKR
KADTTTPATSAVKASSEFSPTFTEKSVALPPIKENMPKNVLPDSQQQYNVETVKVTEQLRHCSEILKEMLAKKHFSYAWPFYNPVDVNALGLHNYYDVVKNPMD
LGTIKEKMDNQEYKDAYSFAADVRLMFMNCYKYNPPDHEVVTMARMLQDVFETHFSKIPIEPVESMPLCYIKTDITETTGRENTNEASSEGNSSDDSEDERVKRLA
KLQEQLKAVHQQLQVLSQVPFRKLNKKKEKSKKEKVNNSNENPRKMCEQMRLKEKSKRNQPKKRKQQFIGLKSEDEDNAKPMNYDEKRQLSLNINKLPG
DKLGRVVHIIQSREPSLSNSNPDEIEIDFETLKASTLRELEKYVSACLRKRPLKPPAKKIMMSKEELHSQKKQELEKRLLDVNNQLNSRKRQTKSDKTQPSKAVENVS
RLSESSSSSESESSSSDLSSSDSSDSESEMFPKFTEVKPNDSPSKEHVKKMKNECILPEGRTGVTQIGYCVQDTTSANTTLVHQTTPSHVMPPNHHQLAFNYQEL
EHLQITVKNISPLQILPPSGDSEQLSNGITVMHPSGDSDTTMLESECQAPVQKDIKIKNADSWKSLGKPVKPSGVMKSSDELFNQFRKAAIEKEVKARTQELIRKHLEQ
NTKELKASQENQRDLGNGLTVESFSNKIQNKCSGEEQKEHPQSSEAQDKSKLWLLKDRDLARPKEQERRRREAMVGTIDMTLQSDIMTMFENNFD

>ZC1_H_SEQID#AA_132
MANDSPAKSLVDIDLSSLRDPAGIFELVEVVGNGTYGQVYKGRHVKTGQLAAIKVMDVTEDEEEEIKLEINMLKKYSHHRNIATYYGAFIKKSPPGHDDQLWLVM
EFCGAGSITDLVKNTKGNTLKEDWIAYISREILRGLAHLHIHVIHRDIKGQNVLLTENAEVKLVDFGVSAQLDRTVGRRNTFIGTPYWMAPEVIACDENPDATYD
YRSDLWSCGITAIEMAEGAPPLCDMHPMRALFLIPRNPPPRLKSKKWSKKFFSFIEGCLVKNYMQRPSTEQLLKHPFIRDQPNERQVRIQLKDHIDRTRKKRGEKDE
TEYEYSGSEEEEEVPEQEGEPSSIVNVPGESTLRRDFLRLQQENKERSEALRRQQLLQEQQLREQBEYKRQLLAERQKRIEQQKEQRRLEBQQRREARRQQBR
EQRRREQEEKRRLEELERRRKEEEERRAEBEKRRVEREQEYIRRQLEEEQRHLEVLQQQLLQEQAMLLECRWREMEBHRQAERLQRQLQQEQAYLLSLQHDHR
RPHFQHSQQPPPQQERSKPSFHAPEPKAHYEPADRAREVEDRFRKTNHSSPEAQSKQTGRVLEPPVPSRSESFSNGNSESVHPALQRPAEPQVQWSHLASLKNNVS
PVSRSHSFSDPSPKFAHHHLRSQDPCPPSRSEVLSQSSDKSEAPDPTQKAWSRSDSDEVPPRVPVRTTSRSPVLSRRDSPLQGSGQNSQAGQRNSTSSIEPRLLWER
VEKLVPRPGSGSSGSSNSGSQPGSHPGSQSGSGERFRVRSSSKSEGSPSQRLENAVKKPEDKKEVFRPLKPAGEVDLTALAKELRAVEDVRPPHKVTDYSSSSEESG
TTDEEDDDVEQEGADESTSGPEDTRAAASLNLSNGETESVKTMIVHDDVESEPAMTPSKEGTLIVRQSTVDQKRASHHESNGFAGRIHLLPDLLQQSHSSSTSTSSS
PSSSQPTPTMSPQTIPQDKLTANETQSASSTLQKHKSSSSFTPFIDPRLLQISPSSGTTVTSVVGFSCDGMRPEAIRQDPTRKGSVVNVNPTNTRPQSDTPEIRKYKKRFN
SEILCAALWGVNLLVGTESGLMLLDRSGQGKVYPLNRRFQQMDVLEGLNVLVTISGKDKLRVYYLSWLRNKILHNDPEVEKKQGWTTVGDLEGCVHYKVV
KYERIKFLVIALKSSVEVYAWAPKPYHKFMAFKSFGELVHKPLLVDLTVEEGQRLKVIYGSCAGFHAVDVDSGSVYDIYLPTHVRKNPHSMIQCSIKPHAIIILPNTD
GMELLVCYEDEGVYVNTYGRITKDVVLQWGEMPTSVAYIRSNQTMGWGEKAIEIRSVETGHLDGVFMHKRAQRLKFLCERNDKVFFASVRSGGSSQVYFMTLG
RTSLLSW